VOL. XLIV
No. 3

PSYCHOLOGICAL REVIEW PUBLICATIONS

WHOLE NO. 199
1933

Psychological Monographs

UNIVERSITY OF IOWA STUDIES IN PSYCHOLOGY

No. XVII

EDITED BY
CHRISTIAN A. RUCKMICK

STUDIES IN EXPERIMENTAL AND THEORETICAL PSYCHOLOGY

PUBLISHED FOR THE AMERICAN PSYCHOLOGICAL ASSOCIATION BY

PSYCHOLOGICAL REVIEW COMPANY
PRINCETON, N. J.
AND ALBANY, N. Y.

PREFACE

This volume contains a group of methodological and experimental papers which have been condensed for publication from the more extended doctorate theses. The theses themselves, together with more extensive protocols and appendices, are on file in the Library of the University of Iowa.

The first paper by Dr. Linder is methodological in scope and proposes some simplifications in a selected group of psychophysical procedures used in the determination of the two-point threshold. It makes a valuable contribution in the sense that some of the more advanced statistical techniques have been brought to bear on psychometric computations. It was completed under the direction of Dr. Tiffin to whom I am indebted for his careful preliminary editing in the direction of a further condensation of the paper. Dr. Murphy offers a contribution to the experimental psychology of the higher cognitive processes. The problem developed out of investigations of reading ability but proceeded toward the creation of some technical devices that served as an attack on the concept.

In Dr. Reger's paper there is significant material on the threshold of extra-auditory sensations as related to the threshold of audibility. The underlying physiological factors are competently discussed and a device is suggested for the use of individuals who are suffering from a considerable loss of hearing. The last paper by Dr. Walker carefully analyzes a number of techniques which he has developed in connection with the photography of eye-movements, especially with material of varying degrees of difficulty and consequent adequacy of comprehension. It is in keeping with the general program of research in our laboratory dealing with these phenomena as related to reading ability but was confined entirely to the functioning of good readers. The last two theses were directed by Dean Carl E. Seashore. Dr. Murphy's thesis came under the direction of the editor.

While the inordinate condensation of these theses placed an

additional burden on the authors, on those responsible for the direction of the respective studies, and finally on the editor, it is a pleasure to note that there was manifest throughout an abundant spirit of coöperation. The editor, therefore, wishes to acknowledge, in a way not easily expressed, his deep appreciation of the attitude taken by these men who in a sense were working with him on a common project. In the last analysis it is an evident example of our combined endeavor which breathes joy throughout the laboratory and inspires each worker in his own work-shop to do the best that in him lies.

THE EDITOR

CONTENTS

Preface	CHRISTIAN A. RUCKMICK	iii
A Statistical Comparison of Psychophysical Methods	FORREST E. LINDER	1
The Rôle of the Concept in Reading Ability	PAUL G. MURPHY	21
The Threshold of Feeling in the Ear in Relation to Artificial Hearing Aids	SCOTT N. REGER	74
The Eye-Movements of Good Readers	ROBERT Y. WALKER	95

A STATISTICAL COMPARISON OF PSYCHOPHYSICAL METHODS [1]

by

FORREST E. LINDER

Psychologists working in psychophysics frequently must choose among available psychometric methods. When such a choice is made the method which gives optimum reliability with minimum experimentation and computation should presumably be used. A large number of papers dealing with a comparison of these methods has been published. Some of these studies [*e.g. Kellogg* (6)] have compared the results of experimentation when the psychological factors are purposely varied. Others [*e.g. Newhall* (10)] have treated the same data by different methods to determine the relative reliability of different computational processes. Still others [*e.g. Urban* (16) and *Thomson* (12)] have attempted to derive general formulæ for expressing the probable magnitude of the sampling errors involved in the various methods.

When the psychological conditions are varied, as in the first approach mentioned above, it is difficult if not impossible to compare the reliabilities of the computational processes *per se*. To accomplish this result we must adopt one of the last two approaches. An inspection of several controversial articles[2] dealing with the third approach indicates that psychologists have not as yet reached agreement upon probable error formulæ in this field. The second approach, namely, that of applying different processes

[1] I wish to make grateful acknowledgment to Dr. Joseph Tiffin for his helpful and critical direction of the thesis and to Dean Carl E. Seashore for originally suggesting the problem.

[2] Among such articles may be mentioned those by *Thomson* given in the bibliography compiled by *Brown* and *Thomson* (3, p. 215). Other contributions to this discussion have been made by *Boring* (2), *Lufkin* (8), *Urban* (17) and in a series of articles by *Culler* (4).

to the same data, would thus seem to be the most promising were it not for the fact that, first, it is difficult to obtain laboratory data in adequate amount for accurate comparison,[3] and secondly, in no case is the theoretically correct value known. We can hardly agree with *Newhall* (10) who takes the most laborious and complicated method as the criterion, for even the values yielded by this method exhibit considerable random fluctuation.

It is the purpose of this paper to describe a new technique for approaching this problem which seems to present certain unique advantages and to give the results of the application of this technique to a comparison of the reliability of four variations of the constant process (*Müller's* method, *Urban's* method, linear interpolation, and interpolation between two points of the integral curve of the normal curve) and the method of limits as developed by *Urban*.[4]

Assumptions and methods of experimentation. We define x_1 as a certain stimulus and assume that there is a constant probability p_1 that x_1 will cause the report of a perception [*Urban*(15)]. In addition, we assume that as x increases from x_1 to x_n, the associated probabilities will also increase, *i.e.,* we assume that p is an increasing function of x. If this function is continuous in the mathematical sense, it is impossible to select any one value of x which is just sufficient to cause a perception. It would seem, due to the physiological and psychological flux of the subject, that the threshold is not constant but varies from instant to instant, and the frequency with which the threshold assumes any value x is a function of x. In this way, the problem of determining the threshold becomes a statistical problem of estimating the central tendency of a frequency distribution. Actually, we are not able to obtain the frequency distribution which we wish to consider. If any stimulus x_1 is presented to the subject, he will report a

[3] *Urban* (15) used seven subjects, *Thomson* (12) used six, and *Fernberger* (5) only two. *Newhall* (10), taking data from several published sources, obtained 100 sets of psychophysical observations, but these obviously involved different psychological conditions.

[4] Descriptions of these methods are given succintly by *Brown* and *Thomson* (3) and *Newhall* (10).

perception, not only when his threshold is at the value x_1, but also when his threshold has any value less than x_1. This is illustrated in Fig. 1.

If the figure illustrates the distribution of the thresholds of a single subject from instant to instant and a certain stimulus value x_1 is presented, then the probability that that stimulus will cause a perception is proportional to the shaded area. For this reason psychophysical data are not represented by frequency-curves, but by *integral* curves of frequency-curves.

It is obvious that if the stimulus is very weak, the probability that it will cause a perception will be almost zero, whereas, when the stimulus increases, the probability that it will cause a perception approaches unity. The exact form of the curve expressing the relation between x and p is probably not the same for all sensory fields, but most psychophysical data seem to be adequately described by the integral of some fairly symmetrical bell-shaped curve.

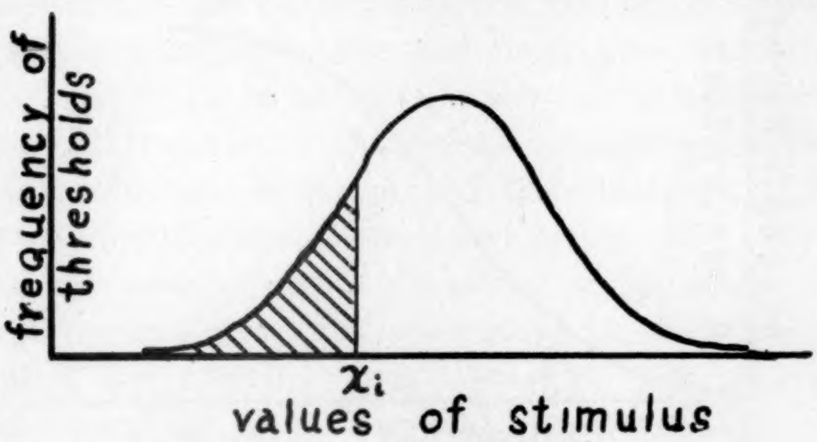

FIG. 1. Hypothetical distribution of thresholds.

Consider a curve:

$$(1) \quad p = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-x^2}{2}} dx$$

which gives a graph of the form illustrated in Fig. 2.

In a convenient manner we consider the origin at the point with an ordinate of $p = 0.5$ which in the derivative curve:

$$(2) \quad y = \frac{1}{\sqrt{2\pi}} \quad e^{\frac{-x^2}{2}}$$

corresponds to the median. Then considering the values of x in units of the standard deviation of (2) we can, by reference to a table of the probability integral, compute values of p for any value of x in which we are interested. If we assume that succes-

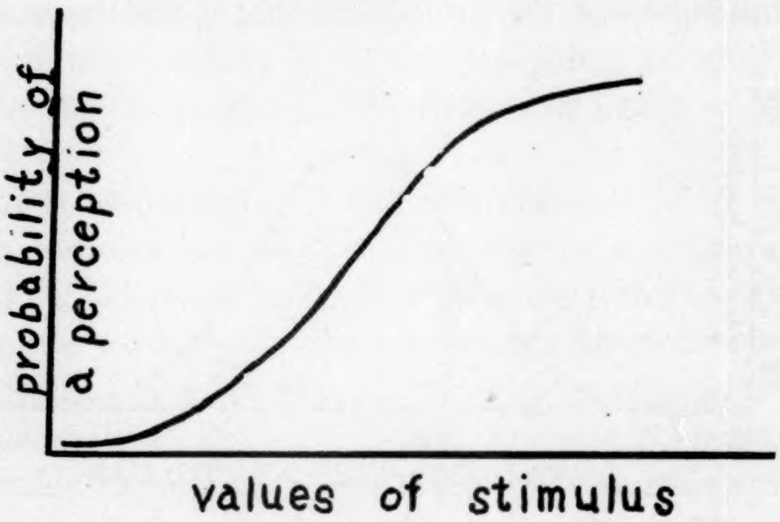

FIG. 2. Integral of frequency curve.

sive values of the threshold are distributed according to (2), then the probability that a perception will be reported for any stimulus value x will be given by (1). An approximation to the central tendency of the distribution of thresholds can be obtained conveniently by interpolating on the integral curve to find the value of x associated with a probability of $p = 0.5$.

The assumption that psychophysical data may be adequately represented by curve (1) has been questioned by *Boring* (1), *Thurstone* (14), *Lufkin* (9), *Thomson* (13), and others. It is beyond the scope of this paper to consider the validity of the use of this curve, commonly called the ϕ (γ) hypothesis. We shall assume its applicability and limit the usefulness of our results

to the analysis of material which is not in serious contradiction to that hypothesis.

We now approximate a laboratory experiment by selecting certain values of the stimulus x. Following the usual laboratory procedure, we can consider about six such values and select them at equal intervals along the scale. Table I gives the values of x selected and the associated theoretical probabilities computed from curve (1).

TABLE I. *Theoretical stimulus values and associated probabilities*

x-values	—2.00	—1.20	—.40	+.40	+1.20	+2.00
Probabilities	.0228	.1151	.3446	.6554	.8849	.9772

In a psychophysical experiment the stimulus value —2.00 would be presented to the subject a certain number of times, with the necessary precautions to eliminate time, space and similar errors, and an empirical approximation to the associated probability obtained. The other stimulus values would be similarly presented. For that one subject we would then have six points on an empirical approximation to the curve in Fig. 2.

We now wish to obtain a number of approximations to the probabilities in Table I in a manner which is free from the psychological variations found in laboratory data. To do this we utilize *Tippett's Random Sampling Numbers* (11). A complete exposition of the use of these numbers is found in the reference given, but since they are not well known among psychologists, it will be serviceable to outline the particular procedure followed.

Tippett's numbers consist of a random selection of 40,000 digits combined to give 10,000 four-digit numbers. The numbers, therefore, range from 0000 to 9999. If we wish, for example, to draw a sample of 50 from a population in which the probability of success is .5000, we can examine any random selection of 50 numbers and tabulate how many of them are equal to, or less than, 4999. It would seem that we should tabulate the number equal to, or less than, 5000, but this would be the case if the lowest possible number was 0001 and not 0000. The term *probability of success* is used in a general sense to indicate the probability of obtaining any defined result. From Table I it is seen that the probability of success when the stimulus value —2.00

is presented is .0228. We get an approximation to this probability, based on 50 observations, by examining the first column of 50 numbers and noting how many are equal to, or less than, .0227. Approximations to the probabilities associated with the other stimulus values are obtained in a similar way. Any set of approximations consisting of each stimulus value with its empirical relative frequency is analogous to the results of the observations made in a laboratory upon a psychological subject. We have avoided, however, all the confusing elements that might occur in laboratory experimentation and have introduced only variations of a random sampling nature. As a result, the data obtained are ideal for comparing the reliability of the computational processes and the sampling efficiency of the various psychophysical procedures.

While the work of drawing the samples is laborious, it is not as time-consuming as laboratory experimentation and consequently the observations were extended to include 100 sets of data.[5] Each set of data consists of 50 observations at each of six stimulus values. As it was desired not only to compare certain psychophysical procedures, but also to study the effect of concentrating the experimental observations on fewer than six points, 100 additional sets of observations were taken. Each set of this group consisted of 150 observations at each of the three stimulus values —0.40, +0.40, +1.20.

The usefulness of the above method of constructing data would be destroyed if there were any correlation between the obtained relative frequencies for the different probabilities. To avoid this possibility, numbers were read from *Tippett's* numbers only once in the same way. Since *Tippett's* numbers contain only 10,000 numbers, and we desired 70,000, it was necessary to go through the book in seven different ways. Such variations in the manner of taking the observations are made by first reading the numbers in vertical columns, then reading horizontally in rows, then taking

[5] **More complete discussion of various points throughout this article is given in typewritten copies of a thesis by *Linder* (7) deposited in the Library of the University of Iowa. These copies include an appendix which contains the data obtained by sampling and values computed by all psychophysical procedures considered.**

the last two digits of each column with the first two digits of the next column, *etc.* In this way every obtained relative frequency was based on a unique group of numbers. Each relative frequency in a set of data and all 200 sets of data are independent in the probability sense.

Before proceeding with the comparison of the various psychophysical procedures by the use of these data it is necessary to be satisfied that they are strictly random. This can be tested in the following manner. For each theoretical probability, by sampling we obtain a set of frequencies. These frequencies form a distribution which can be compared to the expected frequency distribution obtained by expanding the binomial $m(p+q)^s$, where p is the probability of successes, $q = 1 - p$ is the probability of failure, s is the number of observations in each sample and m is the number of samples. The expansion of the given binomial is a laborious task and therefore it was considered sufficient to compare the means and standard deviations of the obtained and expected distributions. If these differ only by an amount such as might be expected under conditions of random sampling, we shall then conclude that our samples are random. For the theoretical distributions the necessary values are given by the formulæ:

$$\text{mean} = sp$$
$$\sigma^2 = spq$$

These theoretical values are compared with the obtained values in Tables II and III.

TABLE II. *Comparison of the mean and standard deviation of frequency of successes in 200 samples of 50 with their expected values*

	Expected	Obtained	Difference	S.E.	Diff./S.E.
p=.0228 and .9772					
Mean	1.135	1.130	.005	.082	.06
S.D.	1.053	1.159	.106	.058	1.83
p=.1151 and .8849					
Mean	5.750	5.675	.075	.159	.47
S.D.	2.256	2.243	.013	.112	.11
p=.3446 and .6554					
Mean	17.225	17.365	.140	.242	.58
S.D.	3.359	3.421	.062	.171	.36

TABLE III. *Comparison of mean and standard deviation of frequency of successes in samples of 150 with their expected values*

	Expected	Obtained	Difference	S.E.	Diff./S.E.
p=.8849					
Mean	132.720	133.030	.310	.405	.77
S.D.	3.910	4.048	.138	.286	.48
p=.3446 and .6554					
Mean	51.675	51.460	.215	.406	.53
S.D.	5.820	5.742	.162	.287	.56

In the second column of each table are given the expected values and in the third column the obtained values. Column four shows the difference between the expected and the obtained values, and column five gives the standard error of the obtained values. For the mean, the standard error is computed by the formula $\text{S.E.} = \frac{\sigma}{\sqrt{N}}$, and for the standard deviation, by the formula $\text{S.E.} = \frac{\sigma}{\sqrt{2N}}$. The value of N is 200 except for the first part of Table III, where $N = 100$. Dividing the difference by the standard error, we get a ratio in column six which can be interpreted to indicate whether or not the difference is greater than one which would occur under random sampling. Generally the ratios are converted to an expression that there are so "many chances in a hundred" that the observed difference would occur with random sampling. This is done by reference to a table of the normal probability integral. With such a procedure the difference is generally not considered to be due to other than random factors, unless the ratio is greater than 2.5. In our case the distributions under consideration are so asymmetrical that reference to the probability integral table is not particularly useful. The ratios given in Tables II and III, however, are so much less than 2.5, that there is little doubt that the discrepancies between the observed and the expected distributions can be attributed to random fluctuations.

We conclude, therefore, that the data collected are typical random samples and we shall proceed to use these data in the discussion of the psychophysical procedures under consideration.

Methods for computing the median threshold. Before pro-

ceeding to the analysis of the results obtained by the various methods it is desirable to define and describe briefly each one considered. The nature of the data obtained depends upon the design of the experiment and the instructions given to the subject. The only methods considered here are those in which certain fixed stimulus values are presented a number of times and the subject is instructed to make a certain kind of report. This excludes such methods of reproduction as the *method of average error*, or the *method of equal-appearing intervals* (**3**, p. 46).

Considering such restrictions, we may say that the usual method of procedure in determining an absolute threshold is to present a certain value of the stimulus to the subject and ask him to report whether or not it is perceived. Several stimulus values are presented a certain number of times and the final data appear in the form of relative frequencies of perceptions, each of which is associated with a certain stimulus value. Then, by some method of computation, the stimulus value associated with the probability $p = 0.5$ is calculated. That is, the stimulus value which is perceived by the subject as often as not, is defined as the absolute threshold. By this definition such a threshold is essentially a *median*. A median threshold can be computed in many ways, but we shall consider only the more customary methods. These may be called: (1) *Urban's* method, (2) *Müller's* method, (3) linear interpolation, and (4) interpolation by the integral curve of the normal curve. Detailed descriptions of these methods are given in several published sources[6] and will not be repeated here.

Comparison of results. Application of each of these methods was made to data secured from *Tippett's* random numbers. The results are shown in Fig. 3. First we may review the method of approach. By the process described we constructed a theoretical population which we assume behaves in a manner similar to a subject under experimental conditions. We take observations at several points of this population just as if we presented various stimulus values to a subject. In the laboratory we obtain certain

[6] *Brown* and *Thomson* (3) and *Newhall* (**10**).

relative frequencies of responses corresponding to each stimulus value and these are mathematically the same sort of relative frequencies that have been obtained here. Thus we have obtained a series of stimulus values and corresponding relative frequencies, which form an analogy to a complete set of psychophysical data. We then take 100 samples and treat these 100 sets of data by the various psychophysical procedures referred to above.

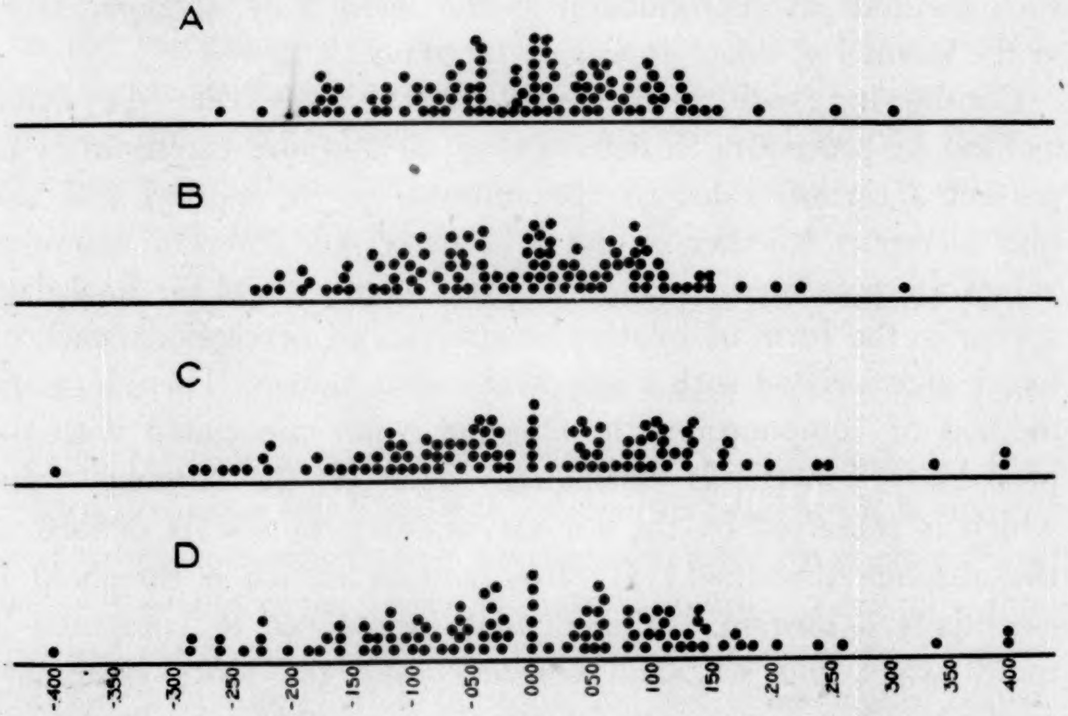

FIG. 3. Distributions of median thresholds: (A) *Urban's* method, (B) *Müller's* method, (C) linear interpolation between two points with 50 observations on each point, (D) interpolation by integral normal curve between two points with 50 observations on each point.

Since we know the characteristics of our population, we are able to tell which psychophysical procedure gives results nearest to this correct value. The theoretical value of the threshold is zero and we shall, accordingly, evaluate the psychophysical procedures by considering which of them gives results nearest to this value. This can be done by considering the nature of the distributions in Fig. 3. If any psychophysical procedure were perfectly accurate, then all of the 100 results obtained would be concentrated at the zero abscissa. Since each method is subject to sampling errors, this is not the case. Obviously the method which concentrates its results most around the zero abscissa may

be considered the most reliable method. Sampling errors of the psychophysical procedures can, therefore, be compared by considering the mean and standard deviations of given distributions. The means, standard deviations, and their standard errors of the four distributions in Fig. 3 are given in Table IV below.

TABLE IV. *Means, standard deviations, and standard errors of distributions of median thresholds*

Method	Mean	S.D.	S.E. Mean	S.E. of S.D.
Urban	—.0106	.1058	.0106	.0075
Müller	—.0087	.1087	.0109	.0077
Linear interp.	—.0073	.1418	.0142	.0100
Normal curve interp.	—.0081	.1467	.0147	.0104

All of the means deviate from the true value, zero, but by not more than one S.E., so we are not able to demonstrate that any method is superior to another when we compare the mean values of a large number of samples. The more critical test is in the comparison of the standard deviations, for this indicates how much variation may be expected if the given method was applied many times to the same situation. We can thus see from Table IV that the standard deviations of the interpolation methods are somewhat larger than the standard deviations of *Urban's* and *Müller's* methods. The differences of these standard deviations must be tested for their significance. The computations on which these tests are made are given in Table V below.

TABLE V. *Comparison of standard deviations of distributions of median thresholds*

Methods compared	Difference of S.D.	S.E. of difference	Diff./S.E. diff.
Urban–Müller	.0029	.0107	0.27
Urban–Linear interp.	.0360	.0125	2.88
Urban–Normal curve interp...	.0409	.0128	3.20
Müller–Linear interp.	.0329	.0126	2.61
Müller–Normal curve interp...	.0380	.0129	2.94
Lin. interp.–Nor. cur. interp...	.0049	.0144	0.34

The S.E. of Difference in Table V are computed from the ordinary formula

$$(3) \quad \sigma_{\text{diff.}} = \sqrt{\sigma_1^2 + \sigma_2^2}$$

However, when there is a correlation between the measures composing the compared distributions, it is necessary to make an

adjustment for this. The formula for the S.E. of difference then becomes (18, p. 210–11)

$$(4) \quad \sigma_{diff.} = \sqrt{\sigma_1^2 + \sigma_2^2 - 2r\,\sigma_1\,\sigma_2}$$

Whenever the correlation between the two distributions is positive, then formula (4) gives a smaller value than formula (3). Consequently, if any difference can be shown to be significant by formula (3), it is not necessary to consider the effect of the correlation as in (4). To do this would only establish to a higher degree of certainty a result which was already established to a sufficiently high degree of certainty.

In statistical practice it is customary to consider a result as established if it is of such a magnitude as would occur by chance with a probability of less than $p = .01$. This is equivalent to saying that the $\frac{\text{Diff.}}{\text{S.E.}_{diff.}}$ given in Table V must be as large or larger than 2.50 (approx.). This criterion of significance is satisfied in all but two cases and these two cases must be examined further by computing with formula (4). These computations are given in Table VI.

TABLE VI. *Comparison of standard deviations of distributions of median thresholds*

Methods compared	r	Difference of S.D.	S.E. of difference	Diff./S.E. diff.
Urban–Müller	.925	.0029	.0028	1.04
Lin. interp.–Nor. c. int.	.988	.0049	.0025	1.95

Both of the values in the last column of this table fall below our criterion of 2.50 and, therefore, the differences compared can not be said to be significant. Even considering the high correlation, the difference between these methods is so small, that the analysis of these 100 samples has failed to demonstrate its significance. This, of course, indicates that the method, adopted by *Urban* of weighting each relative frequency inversely as its variance, adds no more than a negligible amount to the reliability of the method. Similarly our analysis has not indicated any significant difference in the reliabilities of the method of linear interpolation and the method of interpolation by the integral curve

of the normal curve. The results of Table V show, however, that *Urban's* method and *Müller's* method are significantly better than either linear interpolation or interpolation by normal curve.

Thus it is seen that the result of the computation of the above standard errors confirms what is immediately apparent from Fig. 3. Distributions A and B look practically the same and distributions C and D are similar to each other, but either A or B is obviously different from C or D. Nevertheless, while we have shown, for example, that *Urban's* method is better than linear interpolation, it is easily seen from Fig. 3 (A & C) that this difference is not great. We have been able to establish the superiority of one method only by a rather intensive sampling study and for many purposes the small added reliability of *Urban's* method may not be worth the extra amount of work involved.

This introduces us to a consideration of a common criticism of the method of linear interpolation, namely, that it does not use all the data (3, p. 60). This, of course, applies also to the method of interpolation by the integral curve of the normal curve. In drawing the first group of samples, we took 50 observations at six different points, making a total of 300 observations for each sample. In the methods of *Urban* and *Müller,* all 300 of these observations were utilized. The results of the interpolation methods are really based therefore on only 100 observations and not 300 as are the more complicated methods. However, if the laboratory design of an experiment permitted taking a certain number of observations, it would not be necessary to discard most of these, but they could be taken at more favorable places. Thus, if time permitted the experimenter to take 300 observations and he wished to use a method of linear interpolation, it would not be necessary to take 50 observations at six different points, but all 300 observations could be concentrated on two points which subtend the value in which he is interested.

To investigate the effect of this concentration of observations on a few points, additional samples were drawn. The obtained relative frequencies are based on 150 observations at each of the two stimulus values which subtend the theoretical median threshold. For each of these 100 samples, we then computed the

threshold by linear interpolation. Their distribution is shown in Fig. 4 (A). The distribution of the values obtained by fitting the normal curve to the two points is shown in Fig. 4 (B).

Each item of the distributions of Fig. 4 is computed from data composed of 300 observations. These distributions can now be compared with the distributions of results by *Urban's* method and *Müller's* method shown in Fig. 3 (A & B). Again we shall compare the reliability of the methods by comparing the means and

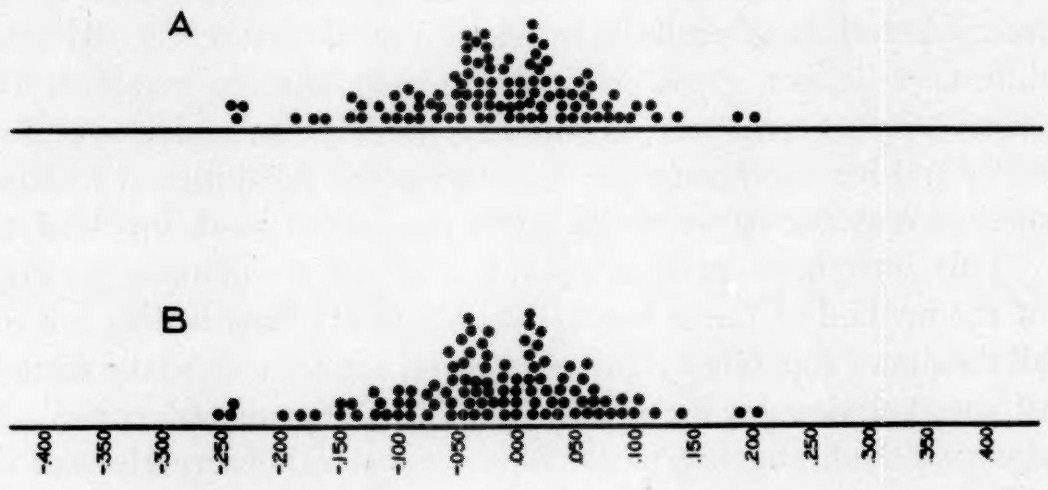

Fig. 4. Distribution of median thresholds: (A) linear interpolation between two points with 150 observations on each point, (B) interpolation by integral normal curve between two points with 150 observations on each point.

standard deviations of the distributions. These values for the distributions of Fig. 4 are given in Table VII.

Table VII. ***Means, standard deviations and standard errors of distributions of median thresholds***

Methods compared	Mean	S.D.	S.E. Mean	S.E. of S.D.
Linear interp.	—.0219	.0806	.0081	.0057
Nor. cur. interp......	—.0208	.0849	.0085	.0060

Let us now compare the distributions described in Table VII with the first two distributions described in Table IV. We shall be particularly interested to see whether the values of one distribution deviate from the theoretically correct value, zero, more than the values of another of the distributions. To do this let us compare their standard deviations. These comparisons and the

computations with which we test the significance of the differences appear in Table VIII.

TABLE VIII. *Comparison of standard deviations of distributions of median thresholds*

Methods compared	Difference of S.D.	S.E. of difference	Diff./S.E. diff.
Urban–Linear interp.	.0252	.0094	2.68
Urban–Normal curve interp...	.0209	.0096	2.18
Müller–Linear interp.	.0281	.0095	2.96
Müller–Normal curve interp...	.0238	.0097	2.45

By noting the values of the Diff./S.E. of diff. which appear in column three of Table VIII we are able to establish that the method of linear interpolation is superior to *Urban's* or *Müller's* method. The values of the Diff./S.D. of diff., in comparing interpolation by the integral curve of the normal curve, are not quite large enough to satisfy our criterion of significance. They differ from 2.50 only by small amounts and in the case of the lowest value of the $\frac{\text{Diff.}}{\text{S.E.}_{\text{diff.}}}$ there are less than two chances in 100 that the observed difference was due to random sampling fluctuations. This leads to the conclusion that the methods used to compute the values of the distributions in Fig. 4 are more reliable than *Urban's* or *Müller's* method.

The results of this analysis make it apparent that if an experimenter has time to make only a certain number of laboratory observations, it will be more efficient to concentrate those observations on a few selected stimulus values and use a simple interpolation method of computation, rather than spread the observations over a number of points and compute the threshold by a more laborious curve fitting process. The argument which we have used to lead to the above conclusion is open to one serious criticism. In the method of sampling which we have used, we have been able to select very easily the points at which we wished to take our observations. Because of this, we have been able to make them subtend in a most convenient manner the value for which we later made our interpolation. With a psychological subject this is not so easy because one is unable to know beforehand what two stimulus values to present in order to obtain one

relative frequency on either side of the relative frequency 0.5. Consequently it may be argued that the experimenter is compelled to scatter his observations over a number of points. But the advantage of concentrating the observations is so great, as may be seen by comparing Fig. 4 with Fig. 3, that the most effective experimental procedure would seem to consist in taking a few preliminary observations to determine the approximate

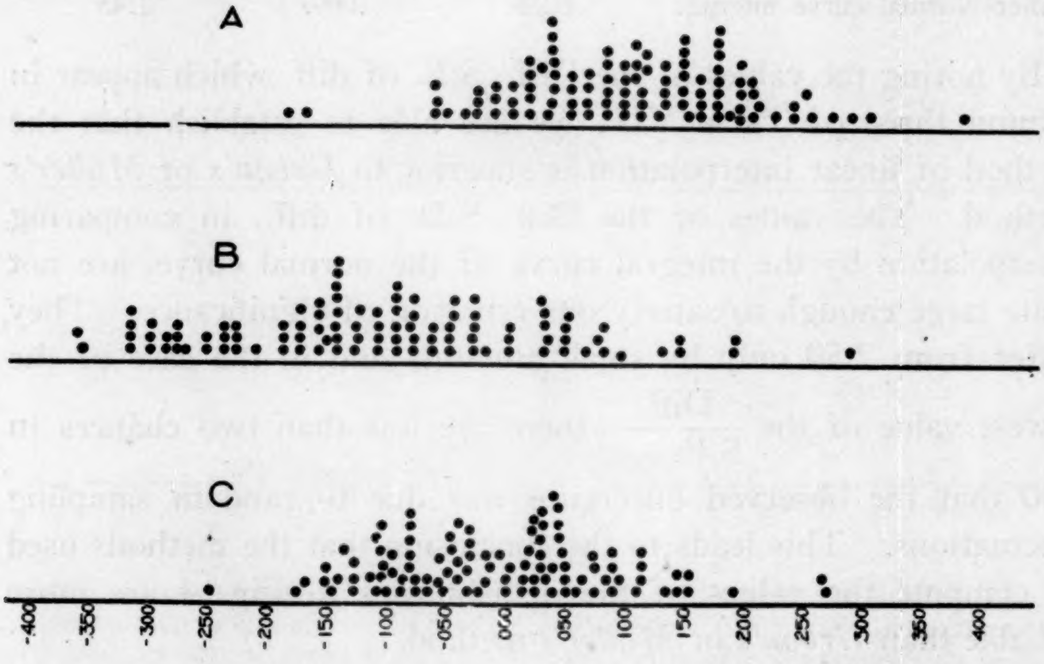

FIG. 5. Distribution by method of limits: (A) mean observed P, (B) mean observed Q, (C) average mean observed P and Q.

value of the threshold and then in concentrating the remaining observations on two points which subtend this approximate value.

The method of limits: A similar comparison was made between each of the methods dealt with in the preceding section and the method of limits described by *Urban* (15). A threshold by the method of limits may be computed in two ways: one is based directly on the observations and the other is computed indirectly from the formulæ developed by *Urban* (15). Detailed instructions for the computation of thresholds by these two variations of the method of limits are given by *Brown and Thomson* (3). Both of these methods of computation were applied to the data previously obtained. The results of the first method, in which

the thresholds were computed directly, are given in Fig. 5. In this figure, A is the distribution of *just perceptible* stimuli (referred to as P), B is the distribution of *just imperceptible* stimuli (referred to as Q), and C the distribution of the means obtained by averaging pairs from A and B.

The results of the second method, in which the thresholds were computed from the formulæ already referred to (4) are given in Fig. 6. A, B, and C in Fig. 6 refer to the same situations as in Fig. 5.

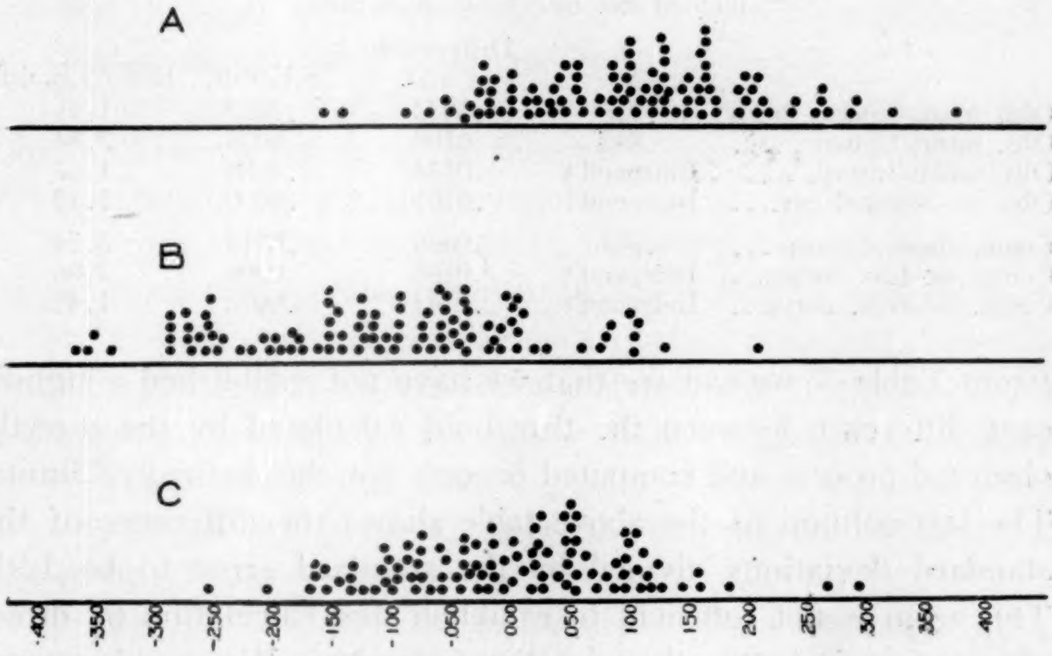

FIG. 6. Distribution by method of limits: (A) mean computed P, (B) mean computed Q, (C) average mean computed P and Q.

The distributions given in Fig. 5 and Fig. 6 are described in Table IX below.

TABLE IX. *Means, standard deviations, and standard errors of distribution of results by method of limits*

	Mean	S.D.	S.E. of Mean	S.E. of S.D.
Mean observed *P*....	.0975	.1006	.0101	.0071
Mean observed *Q*....	—.1060	.1335	.0134	.0094
Distribution of ave...	—.0057	.0950	.0095	.0067
Mean computed *P*....	.0998	.1020	.0102	.0072
Mean computed *Q*....	—.1071	.1255	.0126	.0089
Distribution of ave...	—.0033	.0992	.0099	.0070

We are interested primarily in the distribution of the averages and shall compare these distributions with the distributions

obtained by *Urban's* curve fitting method. We shall also compare them to the distributions obtained by concentrating the 300 observations on two points and computing the median by linear interpolation and interpolation by the normal curve. These comparisons are given in Table X. Since all the means considered are within one S.E. of the mean of the expected values, we compare only the standard deviations.

TABLE X. *Comparison of standard deviations of distributions of results of method of limits with results by Urban's curve fitting method and interpolation methods*

	r	Difference of S.D.	S.E. diff.	Diff./S.E. diff.
Obs. mean–Comp. mean	.928	.0042	.0022	1.91
Obs. mean–Urban	.883	.0108	.0038	2.84
Obs. mean–Interp.	Independ't	.0144	.0088	1.62
Obs. m.–Normal cur....	Independ't	.0101	.0090	1.12
Comp. mean–Urban ...	.967	.0046	.0013	3.54
Comp. m.–Lin. interp...	Independ't	.0186	.0090	2.06
Comp. m.–Nor. curve...	Independ't	.0143	.0092	1.48

From Table X we can see that we have not established a significant difference between the threshold calculated by the directly observed process and computed process for the method of limits. The last column of the above table shows the difference of the standard deviations divided by the standard error to be 1.91. This ratio is not sufficient to establish that calculation by direct observation is better than by the formulæ. We see, however, from the second and fifth row of the table that either the observed mean or the computed mean is more reliable than the values computed by *Urban's* curve fitting method. The difference is not great and we have been able to establish its significance only because of the high correlation between the results. Nevertheless, the apparent superiority of the method of limits can hardly be attributed to chance.

Comparing the method of limits to linear interpolation and interpolation by normal curve when the 300 observations are concentrated at two points, we find the interpolation methods apparently better. Although in neither case have we established that this difference is of a magnitude which would not possibly be attributed

to chance, we can say that it is quite probable that an interpolation procedure, where all the observations are concentrated on a few selected points, is the best procedure for handling psychophysical data of this form. The method of limits is apparently inferior to this, but superior to *Urban's* curve fitting method.

Summary of Conclusions.

1. There is no significant difference between *Urban's* method and *Müller's* method when computing a median threshold. In other words, *Urban's* method of weighting each relative frequency inversely as its variance adds no more than a negligible amount to the reliability of the method.

2. There is no significant difference between linear interpolation and interpolation by the integral curve of the normal curve when computing a median threshold.

3. *Urban's* method and *Müller's* method are more reliable than either linear interpolation or interpolation by the integral curve of the normal curve in computing a median threshold, when the data are obtained at several points and only data at the two points subtending the threshold are used for the interpolation methods.

4. However, when all observations are taken on a few points which subtend the threshold value, the interpolation methods are the most reliable.

5. There is no significant difference between the reliability of thresholds calculated by the directly observed procedure and the computed procedure for the method of limits.

6. The method of limits is more reliable than *Urban's* method.

7. Interpolation methods, when all observations are concentrated on a few points, seem to be more reliable than the method of limits.

8. In all cases in general the most efficient experimental procedure consists in taking a few preliminary observations to determine the approximate value of the threshold and then in concentrating the remaining observations on two points which subtend this approximation value. The threshold can thus be computed by simple linear interpolation or through interpolation by the integral curve of the normal curve.

Bibliography

1. Boring, E. G. The logic of the normal law of error in mental measurement, *Amer. J. Psychol.*, 31, 1920, 1–33.
2. Boring, E. G. The number of observations on which a limen may be based, *Amer. J. Psychol.*, 27, 1916, 315–19.
3. Brown, W., and Thomson, G. H. Essentials of mental measurement, 3rd ed., Cambridge Univ. Press, 1925.
4. Culler, E. A. Studies in psychometric theory. *Psychol. Monog.*, 35, 1926, No. 2, 85–90.
5. Fernberger, S. W. On the relation of the methods of just perceptible differences and constant stimuli. *Psychol. Monog.*, 14, 1912–13, No. 4, 1–81.
6. Kellogg, W. N. An experimental comparison of psychophysical methods, *Arch. Psychol.*, 17, 1929, No. 106.
7. Linder, F. E. A statistical comparison of psychophysical methods (Ph.D. thesis), Univ. of Iowa, 1932.
8. Lufkin, H. M. The accuracy of the method of constant stimuli, *Amer. J. Psychol.*, 38, 1927, 666–67.
9. Lufkin, H. M. The best fitting frequency function for Urban's lifted-weight results, *Amer. J. Psychol.*, 40, 1928, 75–82.
10. Newhall, S. M. Linear interpolation *vs.* the constant process, *Amer. J. Psychol.*, 38, 1927, 390–402.
11. Pearson, K. (Ed.) Tracts for computers, No. XV, Cambridge Univ. Press, 1927.
12. Thomson, G. H. A comparison of psychophysical methods, *Brit. J. Psychol.*, 5, 1912–13, 203–241.
13. Thomson, G. H. Fitting of frequency functions to Urban's lifted-weight results, *Amer. J. Psychol.*, 41, 1929, 70–82.
14. Thurstone, L. L. The $\phi(\gamma)$ hypothesis, *J. Exper. Psychol.*, 11, 1928, 293–305.
15. Urban, F. M. The application of statistical method to the problems of psychophysics, *Univ. of Penn. Stud. in Psychol.*, 1908.
16. Urban, F. M. Die psychophysischen Maasmethoden, *Arch. f. d. ges. Psychol.*, 15, 1909, 261–355; 16, 1909, 168–227.
17. Urban, F. M. Method of constant stimuli and its generalizations, *Psychol. Rev.*, 17, 1910, 229–59.
18. Yule, G. U. An introduction to the theory of statistics, C. Griffin, London, 1919.

THE RÔLE OF THE CONCEPT IN READING ABILITY [1]

by

PAUL G. MURPHY

I. Introduction. One of the most fruitful fields of research in the psychology of reading has been that concerned with the analysis of specific psychological functions involved in the reading process. A cursory perusal of the literature reveals the fact, however, that the great majority of such investigations have dealt with the mechanical phases of the process, such as typographical factors, eye-movements, rate of reading, and related functions, leaving more or less untouched such mental processes as reasoning, conception, and understanding. The present investigation was undertaken in the hope of shedding some light on at least one of these untouched areas, *i.e.,* the rôle played by the concept in reading ability.

Studies of vocabulary have approached the problem in a general way, and are almost unanimous in their postulation of a high correlation between vocabulary skill and reading efficiency. Setting out from this point, the purpose of the present study was to make a more intensive study of the specific nature of this relationship.[2]

II. Historical. Experimental studies of the concept may, in general, be classified under two heads: first, studies dealing with the psychological nature of the concept, and second, studies inquiring into the manner in which concepts are formulated.

[1] Grateful acknowledgment is made to Dr. Christian A. Ruckmick, who directed the investigation, and to Dean C. E. Seashore, who proposed it.

[2] By way of anticipating possible criticism the investigator wishes to put himself on record at the outset as cognizant of the fact that, not concepts in their pure form, but the expressions of such concepts in various kinds of overt behavior constitute the objects of study in this investigation. This limitation is characteristic, however, to a greater or less degree, of all experimental studies of the higher mental processes; consequently, while its force may seem particularly evident in the following pages, the shortcoming is not peculiar to this investigation.

Studies of the psychological nature of the concept. *Ribot* (28), *Kakise* (22), *Jacobson* (19), *Bagley* (2), *Binet* (4), *Schwiete* (31), *Aveling* (1), *Moore* (24), *Fisher* (14), and *Ogden* (25) are among those who have made contributions to this phase of the problem. Two distinct points of view are discernible in the writings of these investigators. One group, including *Ribot, Binet, Aveling, Moore,* and *Ogden* is inclined to deny the dependence of the concept upon any such structural element as the image. They profess rather to have discovered an independent meaning or thought process which accounts for the accruement of significance to the concept and the image is regarded as only an incidental by-product of this process. The second group, composed for the most part of investigators who have conducted their studies under the influence of a structural type of psychology, insist upon the fundamental importance of imagery for concept-meanings.

While he does not take the extreme position of the " imageless thought " school, *Ribot* comes to the highly important conclusion that meaning quite frequently has no representation in consciousness. He contends rather that " this unconscious substratum (the unconscious mind), through organized and potential knowledge, gives . . . an actual denotation to the word,—like harmonics superadded to the fundamental note " (28). A still more important result of this early contribution was the discovery of three conceptual types, analogous to imaginal or ideational types. They are the concrete, wherein the word nearly always evokes an image; the auditory, wherein the auditory image of the word predominates; and the visual typographic, wherein the word is visualized. *Binet,* as a result of various experiments upon his two daughters, concluded that it is the intention or direction of consciousness and not the image itself that is of importance for the development of concept-meanings. Incidentally we may discern in this statement a foreshadowing of the concept of *Aufgabe,* which was developed a short time later by *Watt* and *Ach* of the Würzburg school. *Aveling,* endeavoring to eliminate the uncontrolled factors inherent in the use of familiar words, arranged his experiment in such a way as to induce associations between various meanings and certain nonsense words. He then investigated

the functioning of these words with their acquired meanings in consciousness. He is emphatic in his assertion that universal meanings may appear in consciousness as pure concepts and that the sole contents of consciousness essential to thought are concepts, which are not reducible to sensory elements or complexes. Sensory elements to him are merely by-products of the concept. *Moore* takes essentially the same stand as a result of his investigation of the process of abstraction. He says, " The final product of abstraction, that which is perceived as common to many groups, is essentially a concept distinct from imagery and feeling. It is not an elementary concept, but represents the assimilation of that which is perceived by the sense to a more or less complex mental category, or perhaps to several such categories. The mental category may be regarded as the result of past experience " (24). *Ogden* does not specifically deny the importance of imagery for concept-meanings. He is disposed, however, to relegate it to a minor position, emphasizing the greater importance of certain more strictly functional contents such as the act of relating, and a " dispositional matrix ", which directs this relational activity.

On the other hand, the investigations of *Jacobson, Kakise,* and *Bagley,* all of which were carried out in a predominantly structural atmosphere, are unanimous in their contention that the ultimate constituents of the concept are images and certain feelings, and that the postulation of an additional thought element or process is superfluous. Using letters, words, and sentences as experimental materials, *Jacobson* asked his *O*s to give introspective reports of the processes accompanying the accruement of meaning to these various stimuli. The *O*s were especially instructed to report meaning apart from conscious process. He concluded that meaning comes as a result of the appearance of images and that the meanings of the stimulus words were carried by visual, auditory, and kinæsthetic processes. He also made the point that " the conscious meanings brought out in this experiment are not the static and logical meanings of definition, but rather particular meanings, part-exemplifications, or what not, touched off under the given instruction by the habitual or momentary disposition of the observer. Logically, the representation of meaning is inade-

quate; psychologically, it is adequate to the demands of the occasion" (19). *Kakise,* using a similar technique, attempted to determine the mental concomitants preceding, succeeding, and accompanying the understanding of various concepts. The ultimate constituents of meaning (or concepts), according to him, are images, a feeling of direction, a feeling of content, pure feeling of concept or meaning, and a pure feeling of recognition or familiarity. *Schwiete* confirms him in his discovery of this feeling of familiarity or concept-feeling. Another significant finding in which these two investigators agree is that the appearance of definite images is largely dependent upon the length of time the *O* contemplates the stimulus before reporting the process. *Bagley's* technique varied from that of *Kakise* and *Jacobson* in that he used mutilated words and sentences and presented the stimuli auditorily rather than visually, as had the two latter experimenters. The results of his investigation appear, however, to uphold the validity of a structural analysis of the concept and concept-meanings. He objects strenuously to *Stout's* exposition of "implicit apprehension", contending rather that "the conscious concomitants within the apperception of auditory symbols are made up of sensational and affective elements" (2). *Fisher,* as a result of her study of the process of generalizing abstraction, also affirms the adequacy of the structural point of view, somewhat modified in the direction of functionalism, to the problem at hand. More specifically, she professes to find no evidence in her data for the existence of *Moore's* imageless "mental categories". Even if they should be granted, she contends, they would have to be regarded, along with *Aveling's* "overknowledge of generality", as functional rather than structural categories.

Studies of the development of the concept. A number of exceedingly well-done experimental investigations of the process of concept formation have been carried on since the beginning of the century. While not agreed in their descriptions of the process, they shed considerable light on this important phenomenon of the mental life. Undoubtedly two of the most thoroughgoing attacks on this problem are those of *Fisher* (14) and *Moore* (24). *Fisher* invented four sets of nonsense figures, each

member of a set corresponding to an individual object and the set as a whole corresponding to a species or concept. Of major importance was her discovery that as the experiment progressed images which were at first definite and concrete became progressively more hazy and attenuated. She also found it characteristic of the process that "those contents which prove to be common to a group of perceptions or images obtain an ascendancy over the other contents and hence come to prevail in consciousness". *Moore* is a little more explicit as to the precise steps involved in the formation of a concept. Exposing groups of nonsense figures in rapid succession, he asked the *O*s to indicate the figures common to each group. He discerns the following major steps in the process of abstraction: "the initiation of the process by the breaking up of the group presented for perception, the perception of the common elements initiated by the breaking up of the group and accomplished by assimilating to known categories the sensations perceived, the retention in memory of the figure perceived as common to the group, and the recognition, either consciously or unconsciously, of the universality of this figure" (**24**).

English's study (**11**) is limited to a more specific phase of the process of concept formation: *i.e.,* the initial or abstracting aspect. He distinguishes between two types of abstraction, true abstraction and associative abstraction. The former, which is the only process deserving of the name abstraction, "is characterized by a definite, though perhaps marginal, intention to consider a certain quality in isolation from any of its particulars: *i.e.,* it involves a definite intention to abstract and generalize" (**11**). Associative abstraction may suffice for the everyday needs of life, but the higher modes of thinking almost invariably involve the process of true abstraction. *Hull* (**18**), attempting to formulate an attack on the problem of concept formation analogous to that made by *Ebbinghaus* on memory, has written a monograph on certain quantitative aspects of the evolution of concepts, dealing particularly with the relative efficiency of various methods of formulating concepts. Among other things he concludes that the combination of the methods of abstract presentations and concrete analysis is more efficient for the formation of concepts than either method

alone, that attracting attention to the common element *in situ* increases the efficiency of the process still further, and that moderate familiarity with each member of the series comprising the class group is more efficient than twice as much familiarity with half as many cases. *Smoke* (32), objecting both to the introspective attack on the problem and to the doctrine of common elements, set up his experiment in such a way as to eliminate both these features. He found that the formulation of a concept on the basis of both positive and negative instances was no more effective than when positive instances only were used. As a matter of fact, some *O*s appeared to find the negative instances a hindrance to conceptual learning. He also found no evidence for the assumption that concept formation proceeds either more or less rapidly when the negative instances are far from fulfilling the conditions of the concept than when they violate only one condition essential to it. *Gengerelli's* study (15) clearly demonstrates the aspect of interference in the evolution of concepts. Upon asking his *O*s to evolve two different sets of concepts from the same material (Chinese characters), he found unquestionable evidence of the increased difficulty of the second set. He accounts for this fact on the basis of interference from the first to the second task. Of particular interest from a genetic point of view, also, is *Piaget's* study (26) of the process of concept formation in children. Drawing his material from the observation of children's conversations in natural life situations, his experiments are free from the artificiality inherent in the studies mentioned above. As a result of his studies of children's concepts of causality, the world, time, and space, he emphasizes particularly the fact that their concepts are the result of the juxtaposition and not of the synthesis of a certain number of elements, that there is no real hierarchy, no real composition among the factors, and that it is only later that they come into synthesis.

Studies of the bearing of the concept upon reading ability. More closely related to the theme of this dissertation are investigations of the relationship between the concept and reading ability. Assuming that vocabulary tests are measures of the individual's conceptual equipment, investigators are practically unanimous in

their postulation of a high correlation between adequacy of concepts and reading ability. *Hilliard* summarizes a number of these studies, concluding that they indicate a " general agreement that vocabulary is an important factor in comprehension, although they approach the problem from different angles " (17). *Hilliard* himself finds that vocabulary ranks second only to intelligence in its influence upon reading comprehension. *Jorgensen* (20) reports a correlation of .769 between scores on parts I and II of the Iowa Silent Reading Test, which deal respectively with paragraph comprehension and word meanings. Holding intelligence constant, *Pressey* (27) made various comparisons between a group of poor readers and a group of good readers. Among other things she found that the better readers possessed significantly larger vocabularies than did the poor readers. *Thorndike* (34), in a penetrating analysis of the psychological factors involved in reading, mentions the wrongness or inadequacy of connections between verbal symbols and their meanings as one of the three factors required to explain errors in comprehension. *Dewey's* extensive study (9) of the cause of reading comprehension difficulties in history revealed that ignorance of word meanings lies at the basis of many such difficulties. Attacking the problem from a slightly different angle, *Beall* (3) demonstrated that those units on certain reading tests which are more difficult are almost invariably those containing the largest number of difficult words. Even more pertinent to the problem at hand is *Bird's* study (5), which concerned itself with the effect of concepts broader than words and word meanings upon comprehension ability. One group of normal school students in educational psychology was asked to read, without preliminary preparation or explanation, a difficult chapter dealing with the nervous system. Another group of the same general ability was requested to read the same material except that this time the assignment was preceded by a careful explanation of the various parts of the chapter and a demonstration of a model of a brain and a preserved human brain. In terms of the present study it may be said that the members of the second group attacked the reading material with broader and more adequate concepts of the topic treated in the reading material than

those of the first group. On a subsequent comprehension test designed to test their understanding of the material read, the experimental group made significantly higher scores than did the control group. *Condit's* study (8), which is practically identical with that of *Bird* except that it was carried on at a lower age level, failed to show such large differences between the two groups. In every case, however, the difference was in favor of the training or "apperception" group, a fact which can hardly be accounted for on the basis of chance.

III. The experimental approach. At least two approaches to the problem suggested themselves at the outset: first, the method of correlation, and second, the method of comparisons between groups representing the extremes of reading ability. The latter was selected as the basis for this investigation for several reasons. First, we felt that this method would throw into relief differences in concepts concomitant with differences in reading ability that might be obscured by the correlational technique due to low relationship between reading efficiency and the other factors in the middle range of ability. Secondly, we desired to make an intensive study of a comparatively small number of cases rather than an extensive statistical study of a large number of cases and it was felt that the method of divergent groups met the demands of such a procedure more adequately than the method of correlation. And thirdly, setting up the problem in such a way as to obtain certain constancies as well as certain discrepancies within various abilities appeared to be a much simpler task by this method.

Observers. Two groups of University of Iowa freshmen representing the extremes of efficiency in reading ability and vocabulary skill were selected to serve as *O*s in this study. One group, hereafter designated as "good readers", was composed of 10 individuals ranking at the 65th percentile or above in comprehension ability and at the 58th percentile or above in vocabulary skill, as these functions are measured by parts I and II, respectively, of the Iowa Silent Reading Test. The second group, hereafter referred to as "poor readers", was made up of 10 individuals ranking at the 36th percentile or below in comprehension ability and at the 40th percentile or below in vocabulary skill.

That the two groups did not represent further extremes of ability in these functions is due to the fact that they were also selected in such a way as to be of approximately equal intelligence at the 50th percentile, as this capacity is measured by the University of Iowa Qualifying Examination, exclusive of the Iowa Silent Reading Test.[3] The correlations between reading ability and intelligence and between vocabulary skill and intelligence are so high that it was not possible to secure individuals who deviated more widely from the 50th percentile in these capacities and yet who were of approximately equal intelligence.[4]

Despite these facts, the investigator was relatively successful in obtaining groups that fulfilled the requirements of the experiment. The difference between the average scores of the two groups on part I of the Iowa Silent Reading Test was 12.5 times the S.E. diff., while the difference between the average scores of the groups on part II of this test was 8.84 times the S.E. diff. Incidentally, this latter difference was further substantiated by the finding that the difference between the average scores of the groups on one part of the Minnesota College Aptitude Test, which was used in a subsequent experiment, was 6.26 times the S.E. diff. On the other hand, the difference between the average scores on the Qualifying Examination, wherein equality was desired, was .04 times the S.E. diff. In view of the fact that 3 times the S.E. diff. constitutes a significant difference, it is obvious that highly significant differences between the two groups were obtained in those capacities wherein differences were sought, while an approximate equality in intelligence was maintained.

It may be safely assumed, therefore, that variations in other capacities which may have been brought to light in the course of this study can not be assigned to differences in intelligence. There may be other factors which were not controlled and to which these variations may be attributed. For the purposes of this investigation, however, it would seem safe to conclude that such variations are functionally correlated in some manner with reading ability.

Systematic considerations. As a preliminary to the selection of the specific techniques to be utilized in analyzing the concepts of the individuals chosen for study, the following systematic issues were considered: (1) what is a concept?, and (2) what comprises an adequate description of a concept?

As *Fisher* has pointed out, two more or less distinct points of view are discernible in discussions concerned with the nature of the concept. On the one hand

[3] The validity of using this examination as a measure of intelligence is indicated by the fact that the correlation between scores on this test and scores on the American Council of Education Psychological Examination is +.80.

[4] The original data of this research, as well as sample copies of the various test techniques utilized, are contained in copies of the dissertation which have been deposited in the Library of the University of Iowa.

we find the "cognitive" theories, "which attempt to envisage the concept in ideational terms"(14). This point of view is characteristic of the early philosophical and logical discussions of the concept, as well as the later writings of *Wundt, Störring,* and *Lipps.* With the advent of a more dynamic type of psychology, "motor" theories of the concept came to be held. These "find the essence of the concept in a motor phenomenon or tendency, with or without a conscious coefficient of kinæsthesis or feeling"(14). This comprises the essential point of *Bolton's* discussion of "meaning as adjustment"(7), of *Wheeler's* statement that "meaning is referable to kinæsthesis"(37), of *Weiss's* description of meaning as response (35), and of the behaviorist's definition of a concept as "a group of responses which have been classified together and labeled by a verbal symbol"(16). *Fisher* also inclines toward the "response" view of the concept.

The concept has also been defined in terms of the third aspect of mental life, feeling. Both *Drever* (10) and Rignano have suggested that the essential feature of the concept is its affective component. The former insists that any analysis of the meaning of words from the cognitive point of view necessarily leads to a logical investigation, and that meaning must be regarded as primarily affective. The latter, who is disposed to explain all forms of reasoning from an affective standpoint, states that "the most familiar common nouns or concepts represent nothing but classes of objects, which may, we repeat, differ extremely among themselves from other points of view, but which are all equivalent in relation to some of our needs or of our affective tendencies. In other words every common noun, every concept is essentially an affective grouping"(29).

In spite of the varying definitions that have been proposed for the concept, practically all experimental studies of this psychological phenomenon have been undertaken from a "cognitive" point of view. It is usually, however, the overt expression of the concept, either in language or behavior, that provides the immediate data for study; nevertheless, the more intangible but supposedly more fundamental cognitive experience which is assumed to underlie this overt behavior, is of ultimate interest to the investigator. The writer has few qualms, then, in admitting that for the purposes of this investigation the concept has been defined, from the angle of content, as an item of knowledge, and from the angle of form or process, as "an object of awareness together with its significance or meaning"(12).

It was with the purpose in mind of providing the study with a systematic basis of procedure that a description of the concept was undertaken. Specifically the question to be answered was, what attributes does the concept possess? When a sensation has been analyzed from the standpoints of attensity, intensity, protensity, and quality, an exhaustive description of it is available. In an analogous fashion we may ask, what are the attributive characteristics that must be taken into consideration in describing the concept?

A survey of the literature revealed no direct efforts at such an analysis. The few clues that were unearthed, however, seemed to point toward the essential validity of the following analysis. In the first place, the traditional dichotomous classification of attributes into *qualitative* (or, in this case, *generic*) and *quantitative* would appear to be applicable. Differences of a *generic* nature are perhaps not so clearly discernible as in the case of sensation; however, such differences as are exemplified in concrete as opposed to abstract concepts appear to be of this order. The varying mental content of concepts might also serve as a basis for generic classification. The applicability of various *quantitative* categories is more obvious. From this point of view, each concept possesses a given

degree of *richness*, completeness, or fullness; it stands in a certain degree of *accuracy* or fidelity to the accepted meaning of the concept (which, for practical purposes, may be said to be contained in the formal dictionary definition of the concept); it possesses a definite degree of *organization;* and it is characterized by a certain amount of *clarity*.

In terms of this scheme, then, the purpose of the investigation was to determine how the concepts of good and poor readers vary, first, generically, and secondly, quantitatively, the latter category subsuming the attributes of richness, organization, clarity and accuracy.

IV. Generic differentiæ: abstract and concrete concepts. The acceptance of the abstract-concrete dichotomy as the most obvious basis for a generic classification of concepts immediately suggested experimentation designed to ascertain whether the concepts of the two groups of readers described above varied significantly in either or both of these categories. Such experimentation must, however, be taken up from a particular point of view, *i.e.*, from the angle of determining whether the two types of concepts vary in accuracy, in richness, in mental content, or in organization. The two following studies, as well as certain of the various quantitative experiments, were set up, therefore, in such a way as to enable us to compare the abstract and concrete concepts of the two types of readers from several rather specific points of view.

Series A: the mental content of concepts. From the point of view of structure, variations in the mental content of concepts may provide a basis for generic differentiation. It was upon this assumption that an introspective study of concepts was undertaken.

Preliminary experimentation. Since the *O*s included in the groups selected for study were completely naïve concerning scientific psychological methods, the usual technique for obtaining introspective reports had to be modified quite radically. After a consideration of various possibilities, the method of the "standardized interview" was adopted, whereby introspections were obtained upon the basis of a series of leading questions put to *O* by *E*. It was for the purpose of providing information necessary for the construction of the interview that this preliminary study was undertaken. Ten *O*s were used in this experiment. Eight were graduate students and two were members of the staff in the

psychology department at the University of Iowa. All possessed some degree of experience in introspective observation.

In an attempt to reproduce the reading situation as nearly as possible, the concepts were presented in a context of reading material composed of four comparatively short selections taken from the fields of science, religion and history. *O* first read through the selection at his normal reading rate. He was then requested to close his eyes, whereupon a duplicate of the original selection, containing certain underlined words and phrases, was placed before him. A blank piece of cardboard was placed over the material in such a way that only the first underlined phrase in the passage was exposed. At a given signal, *O* opened his eyes, fixated the underlined phrase, gave a short definition of the word or phrase in the sense in which it was used in the paragraph and followed immediately with an introspective account of the mental process accompanying the arousal of the concept in consciousness. Variations of this general procedure, including the presentation of words in isolation from any context, and reports on the process accompanying the mere arousal of the concept, apart from the act of defining, were also utilized.

Since the primary purpose of this experiment was to provide information necessary for the construction of a procedure to be used in the main experiment, no further analysis of the introspective report than that exemplified in the " standardized interview " (see appendix) was attempted. The *O*s serving in this study were quite generally agreed that, cognitively, there was little or no difference in the conscious process accompanying the arousal of concepts in context and in isolation from any context. Affectively, the latter situation was attended by a feeling of greater freedom, greater latitude than the former; other than this, however, no major distinctions were reported. For this reason, no attempt was made to reproduce the reading situation in the following study.

Procedure. Twenty *O*s comprising the two experimental groups served in this experiment. All experimentation was carried on in the same room and, as far as possible, under the same conditions of temperature, humidity and illumination.

A series of six isolated terms, three abstract and three concrete, were used as experimental materials. Six white cards, each containing a single term, were placed before *O* and he was instructed as follows:

"There is a single word printed on the reverse side of each of these six cards. You are to turn the cards over, one at a time, and give me a definition for each word. Do not hurry for you will be given plenty of time to formulate each definition. On the other hand, do not deliberate too long on a word. A simple statement of the meaning of the term is all that is desired.

After you have given your definition, I will ask you some questions about what went on in your mind as you were formulating your definition. For the sake of accuracy, answer the questions to the very best of your ability. We are not interested in how good a definition you can give, but only in the mental process whereby you arrive at the definition you do give. You may ask any questions you like about the questions I ask. If any of the terms I use are not clear, ask about them."

Following the definition of each term, *E*, with the aid of the interview, obtained an introspective report from *O*, recording his statements directly on the form containing the interview questions.

Results. To facilitate the drawing of comparisons between the reports of the two types of readers, the results of this experiment have been summarized in Table I. An analysis of 120 introspections, 60 from each group, is represented in this summary. The 60 reports obtained from each group have been further divided into two sets of 30, one set containing introspective reports based on abstract terms and the other including reports upon concrete terms.

The manner in which the table is to be read may be illustrated by a consideration of the first item. Of the instances of visual imagery of the stimulus word reported in the 60 reports of the good readers, 55 per cent were found in the 30 reports dealing with abstract words and 45 per cent in the 30 concerned with concrete terms. Therefore, 10 per cent more items of this type were reported in the abstract than in the concrete reports. This fact is indicated by the +10 per cent in the fourth column. Had the difference been in favor of the concrete category, a —10 per cent would have been indicated here. Of the 18 instances of visual imagery of the stimulus word reported in the 60 reports of the poor readers, 56 per cent were found in the 30 reports

dealing with abstract terms and 44 per cent in those dealing with concrete terms, leaving a difference of +12 per cent in favor of the abstract category. The percentages for the abstract and concrete categories have not been computed in cases where the gross number of items reported was exceptionally small.

While no attempt was made to test the differences indicated in Table I for significance, it is obvious that only a few if any would have met the demands of reliability. The difference between the gross number of items reported by the two types of readers in items 1A, 1B, 2B, and 2D may be true variations. The weight of evidence as a whole, however, is clearly in the direction of structural similarity rather than dissimilarity between the conceptual consciousnesses of the two groups.

The information necessary for a comparison of the imagery of the two groups on the bases of relative distinctness, temporal duration and specificity was also obtained in these reports. In general the imagery of both groups tended to be fairly distinct, although fleeting. From the point of view of specificity, the images tended to be about equally divided between delineations of a general type and those of a more specific nature, exemplified for the most part in memorial images. The deviations of either group from these tendencies were found, however, to be exceedingly slight, which, of course, is entirely in keeping with the general trend of results obtained in this experiment.

The mental content of abstract and concrete concepts. Differences in the mental content of abstract concepts of the two types of readers may be discerned by a comparison of the percentages in columns (2) and (6); a similar study of concrete concepts is afforded by the figures recorded in columns (3) and (7). Probably more significant, though, are the relationships between abstract and concrete concepts expressed by the differences recorded in columns (4) and (8). One of the more striking tendencies revealed by this latter comparison is the predominance of visual imagery of words not included in the definitions in the concrete reports of good readers and in the abstract reports of poor readers. A tendency of the same type and in the same direction is also to be noted in the field of motor imagery. The temp-

TABLE I. *Analysis of the reports made by good and by poor readers*

	Good readers				Poor readers			
	(1) No. items	(2) Abs. (in per cent)	(3) Con. (in per cent)	(4) Diff. between A & C	(5) No. items	(6) Abs. (in per cent)	(7) Con. (in per cent)	(8) Diff. between A & C
1. Visual imagery:								
A. Of stimulus word	29	55	45	+10	18	56	44	+12
B. Of the definition	23	52	48	+4	17	53	47	+6
C. Of other words	21	48	52	—4	27	59	41	+18
D. Of places, objects, people, or events	82	37	63	—26	92	35	65	—30
Total	155				154			
2. Auditory imagery:								
A. Of stimulus word	4				4			
B. Of the definition	0				6			
C. Of other words	1				3			
D. Of places, objects, people, or events	6	50	50	0	11	55	45	+10
Total	11				24			
3. Motor imagery:								
A. Of the stimulus word	44	61	39	+22	44	55	45	+10
B. Of the definition	44	52	48	+4	37	46	54	—8
C. Of other words	24	33	67	—34	24	54	46	+8
Total	112				105			
4. Ideas not involving imagery	45	62	38	+24	52	60	40	+20
5. Memories (excl. memorial images)	11	55	45	+10	13	46	54	—8
6. Confusion while forming def.	35	60	40	+20	32	62	38	+24
7. No confusion	25	36	64	—28	28	36	64	—28
8. Satisfaction with def.	31	42	58	—17	35	54	46	+8
9. Dissatisfaction with def.	29	60	40	+20	25	44	56	—12
10. Many ideas accompanying arousal of concept	45	51	49	+2	41	54	46	+8
11. Few ideas accompanying arousal of concept	15	47	53	—6	19	42	59	—16
12. Other meanings of the concept than the one given also aroused in consciousness	7				1			

tation to see in these facts evidence of the further progress of good readers up the scale of abstract thought is belied, however, by the close similarity of the two groups in practically all the other items dealing with imagery.

Memorial ideas appear to predominate in the abstract reports of good readers and in the concrete reports of poor readers; the significance of this difference is highly questionable, though. In only two other items, those concerned with the relative amount of satisfaction or annoyance accompanying the formulation of the definitions, are differences of any import to be found. Here the good readers profess to be more satisfied with their definitions of concrete words, while the poor readers are more inclined to accept as adequate their definitions of abstract terms. Since, however, the percentages on these items in the abstract and concrete columns do not deviate more than 10 per cent from the 50 per cent point for either type of reader, the existence of any reliable differences here is questionable. From the structural aspect at least, none of the variations revealed in a comparison of the totals appearing in columns (4) and (8) are sufficiently large to be indicative of any striking dissimilarities between the abstract and concrete concepts of good and poor readers.

Series B: association types. A second approach to the problem of generic differences in the concepts of good and poor readers was attempted in a comparative study of the types of responses exhibited by the two groups in a free association test. The assumption underlying the utilization of the free association technique here is that associative types may be diagnostic of certain qualitative differences in concepts. This assumption is not altogether without support. *Bleuler,* for example, says: " Thus in the activity of association there is mirrored the whole psychical essence of the past and of the present, with all their experiences and desires " (6). Concepts seem to be outgrowths of such " experiences and desires ".

Wells (36) has shown that the process of classifying associations by types is fairly reliable, the correlations between the classifications of two sets of results obtained eight months apart averaging approximately .80 for the various categories. He also

found that the utilization of different sets of stimulus words did not produce variations sufficient to obscure real individual differences in association types.

Procedure. Fifty words, selected from the Kent-Rosanoff free association test (30) in such a way as to include an equal number of abstract and concrete terms, served as experimental materials. Using the Whipple tachistoscope as an exposure device, these terms were presented to the twenty *O*s of the experimental groups, with the instruction to " respond each time with the first word that comes into your mind ". The responses thus obtained were classified into six categories, according to a modification of *Jung's* scheme (21). The types of connections holding between stimulus and response in the various classes are described in the following table.

TABLE II. *Types of associative responses*

1. Inner associations...........	contrast, subordination, supraordination, predication.
2. Outer associations...........	contiguity, part-whole, word-completion, syntactic alteration, repetition of stimulus word.
3. Clang associations..........	associations based on similarity of sound.
4. Egocentric associations......	referring to self, inexplicable associations.
5. Perseverative associations...	associations influenced by previous responses.
6. No response	

Results. An analysis of the responses according to the above scheme yielded the following results for the two groups:

TABLE III. *Classification of associative responses to 50 stimuli*

	Ave. no. associations per *O*				
	Good readers (N=10)	per cent	Poor readers (N=10)	per cent	P[5]
1. Inner associations........	30.4	(61)	32.5	(65)	38
2. Outer associations.......	17.0	(34)	14.7	(39)	31
3. Clang associations.......	0	(5.2)	.3	(5.6)	
4. Egocentric associations...	.7		.8		
5. Perseverative associations	1.4		1.4		
6. No response	.5		.3		

The high values recorded for P in the first two categories make it obvious that the two groups of readers did not differ reliably in

[5] P indicates the number of chances in 100 that the various observed differences are due to chance, according to *Fisher's* method (13) for computing the significance of differences between the means of small groups.

the average number either of inner or of outer associations. The small number of examples falling in the last four classes made the computation of P for each category a questionable procedure. It seems doubtful, though, if any significant differences would have been indicated.

A check on the reliability of the above results was afforded by the following analysis of the responses of another group of 14 poor readers on a 25 word free association test used as a diagnostic technique in the reading clinic at the University of Iowa.

TABLE IV. *Classification of associative responses to 25 stimuli*

	Ave. no. associations per *O* (N=14)	per cent
1. Inner associations	19	(76)
2. Outer associations	5	(30)
3. Clang associations	0	(4)
4. Egocentric associations	.57	
5. Perseverative associations	0	
6. No response	.14	

While the proportions indicated for the various types of responses are not identical with those indicated for poor readers in Table III, they are sufficiently similar so that the conclusions drawn in the preceding paragraph appear still to hold. Such discrepancies as do exist can probably be accounted for on the basis of differences in the two sets of stimulus words.

Associative responses to abstract and concrete stimuli. Since the 50 word free association test described above was composed of an equal number of abstract and concrete terms it was possible to continue the study of logical distinctions between the concepts of the two types of readers which was begun in Series A. A reclassification of the data contained in Table III, according to whether the associations were responses to abstract or concrete stimuli, provided the material for the construction of Table V.

TABLE V. *Logical classification of associative responses to 50 stimuli*

	Ave. no. associations per *O*					
	Good readers (N=10)		Poor readers (N=10)		P	P
	Abs. stim.	Con. stim.	Abs. stim.	Con. stim.	Ab.	Con.
1. Inner associations	19.5	10.9	19.9	12.6	72	30
2. Outer associations	4.1	12.9	3.9	10.8	81	22
3. Clang associations	0	0	.2	.1		
4. Egocentric associations	.5	.2	.4	.4		
5. Perseverative associations	.2	1.2	.3	1.1		
6. No response	.3	.1	.1	0		

Here again no differences deserving of consideration are indicated. The division of inner responses among abstract and concrete stimuli is practically the same for both types of readers, as is also the division of outer responses.

Decided differences are apparent, however, when we disregard the " good reader—poor reader " classification and consider the division of inner and outer associations among abstract and concrete stimuli in general. A much larger number of inner responses were made by the *O*s of both groups to abstract than to concrete stimuli, while just the reverse is true of the outer responses. These facts were probably predetermined, though, both by the manner in which the test was constructed and by the scheme utilized for classifying responses. For instance, the fact that at least half the abstract stimuli were adjectives taken together with the fact that all predicative responses were placed in the first category probably accounts for the preponderance of inner responses made to abstract stimuli. In a like manner, the nature of the associative relationships included in class two probably accounts for the large number of outer associations made to concrete stimuli by both groups of readers.

Summary and conclusions. Three approaches to the problem of generic differences in the concepts of good and poor readers were attempted. While the first, an introspective study, indicated three or four possible points of variation, the weight of evidence as a whole pointed clearly in the direction of structural similarity rather than dissimilarity between the conceptual consciousnesses of the two types of readers. Neither did a classification of responses on a free association test reveal any differences worthy of consideration. The third approach was afforded by a comparison of the abstract and concrete concepts of the two types of readers, first, from the angle of mental content, and secondly, from the point of view of associational types. The results of this study only served to support the conclusions of the two previous studies. In so far as these techniques are valid measures of the generic aspects of the concept, then, no such variations have been demonstrated by these studies.

Of course, it is entirely possible that the impotence of this

study to bring to light generic differences between the concepts of the two types of readers is largely a function of the gross nature of the techniques here utilized. *Boring* has pointed out, however, in his *History of Experimental Psychology,* that " research in any science must arise out of available methods ", and it is generally recognized that the methods in this field of investigation are quite coarse. This should not prevent us, however, from pointing out possible points of difference between the two types of readers as well as suggestions for further research.

V. Quantitative differentiæ: richness of the concepts of good and poor readers; Series C: the free association test as a measure of richness. By the richness of a concept we mean, theoretically at least, the number of items of information included in a general notion. From this point of view the most obvious approach to an evaluation of the richness of an individual's concepts would be to determine the number of items of information included in his repertoire of knowledge on specific topics. The excessive demands that such a procedure would make upon both the time and patience of *E* and *O,* however, ruled it out as impractical.

The free association technique was finally utilized as best adapted to the demands of the problem. While the applicability of this procedure is not so apparent as that mentioned in the preceding paragraph, it becomes clear upon consideration that the richness and rapidity of association is conditioned, among other things, upon the richness of concepts. Where little meaning attaches to the stimulus word the responses will be slow and few; where many meanings attend the arousal of the concept in consciousness they will be many and rapid. Support for this assumption was discovered, after the technique had been in use for some time, in a study by *Wiley* (39), wherein he states, " Quickness of response [to stimuli in a free association test] is perhaps one of the best measures available for richness of association ". The fact is recognized, of course, that the capacity for verbalization is also involved here. Again, this would appear to be a shortcoming contingent upon the fact that we are limited, in any such study as this, to a consideration of the linguistic expression of

concepts rather than the concepts as they actually exist in the consciousness of the individual.

Preliminary experimentation. In order to gain some experience in the use of the continuous free association technique, the investigator undertook a preliminary experiment involving 14 good readers and 27 poor readers. The average percentile ranking of the former group on part I of the Iowa Silent Reading Test was 80 as compared to average ranking of 18 for the latter group. On part II of this test the respective average percentile rankings were 77 and 21; on the 1930 Qualifying Examination, 93 and 17.

Ten terms, five abstract and five concrete, were used as experimental materials in this study. These were presented auditorily with the instruction to " respond with all the words that come into your mind during the 15 seconds following the hearing of the word ". The stop-watch was started at *O*'s first response and time was called at the end of the 15 second interval. No effort was made to record the responses, an enumeration of the number of responses being all that was desired in this study.

Table VI summarizes the results of this experiment.[6]

TABLE VI. *Analysis of responses on continuous free association test*

Ave. no. associations per stimulus word		
Good readers (N=14)	Poor readers (N=27)	$\frac{\text{Diff.}}{\text{S.E. diff.}}$
7.79	6.48	2.11

Procedure. The 20 *O*s of the regular experimental groups were utilized here. In order to observe the effect of variations in the stimulus materials upon the subsequent responses, 10 words, 10 short phrases, and 10 paragraphs of medium length, equally divided between abstract and concrete stimuli, were used as experimental materials. Since the visual mode of presentation seemed preferable in certain respects to the auditory method used in the preliminary experiment, the stimuli were exposed on small white

[6] The number of cases utilized here was sufficient to warrant the use of the more familiar formula for computing the significance of the difference. In most of the experiments, however, the fact that only 20 *O*s were used made it necessary to resort of the more conservative P-formula devised by Fisher for use with small groups.

cards, one to a card. As in the introspective study, the cards were presented to *O* face down, and the following instructions were given:

"On the reverse side of each of these cards is one word. You are to turn the cards over, one at a time, and as soon as you see the word each time you are to respond with a series of words indicating all the things that come into your mind during the 30 seconds following. For instance, on seeing the word *horse* you might think, in succession, of a farm, walking, hair, legs, work, *etc.*, in which case, of course, you would respond with these words. Or you might think of entirely different groups of things. The things that come into mind may not seem to be related in any way to the stimulus word, but you will say them anyway. You are to name *all* the things that you think of during the 30 seconds. Do not respond with phrases or sentences, but with a series of single words. The idea is to find out what various things come into your mind upon seeing the words. Let's try a sample or two to make sure that you understand."

In the case of the paragraphs, *O* was instructed to read through each paragraph at his normal reading rate, then lay it aside face down and respond as with the words and phrases. The stop-watch was started at *O*'s first response and time was called at the end of the 30 second interval. A stenographic record of the responses given by *O* was taken down by *E*.

In addition to noting responses in the discrete free association test utilized in Series B, reaction times were also recorded. It was possible, therefore, to incorporate the quantitative results of this study into the present experiment as another check on differences in the relative richness of the concepts of the two types of readers, this time from the point of view of rapidity rather than richness of response.

Reaction times were measured by placing a Cenco Impulse Counter, measuring time intervals in 120's of a second, in circuit with a Whipple tachistoscope which was used in this case merely as an exposure device. The circuit was so arranged that at the instant the stimulus was exposed, the counter was set in motion. When *O* responded, *E* stopped the counter by opening a reaction-key which had been placed in the circuit. The time of each reaction could be computed, then, by subtracting successive readings recorded from the timing device. Of course, *E*'s reaction time entered into each such computation; however, it was assumed that this would be fairly constant for 50 reactions and so would not affect any comparisons that might be made between individuals or between groups.

In order to control illumination the experiment was conducted in a "dark room". A lighted pre-exposure field of constant intensity was provided by a 60-watt lamp hung directly above the

exposure apparatus. The stimulus field was lighted by a frosted lamp which is standard equipment on the tachistoscope.

At the beginning of the series the following instructions were read to *O:*

"There will be exposed here, one at a time, a series of words. Each time you see a word I want you to say aloud the first word that comes into your mind. The word you say needs to bear no particular relationship to the stimulus word. You say the very first word that comes into your mind. Say it as quickly as you can. Do not respond with a phrase or sentence but with a single word only. Do you understand?

Instead of my exposing the words, each time I will give you the signal 'ready' and then, as soon as you feel that your whole attention is fully concentrated on the task, you may pull this cord and the word will be exposed to view. Let's try a few samples to make sure that you understand what you are to do."

Three sample stimuli were exposed and any questions that *O* had to ask about the procedure were fully answered before proceeding with the series.

Results. The results obtained in these two experiments are summarized in Table VII. Due to the fact that a few exceptionally long reaction times unduly weighted the average reaction times of certain individuals, the median as well as the mean of the 50 time intervals obtained for each individual on the discrete association test was computed.

TABLE VII. *Analysis of responses on free association tests*

	Continuous free association test		
	Ave. no. associations per stimulus		
	Good readers (N=10)	Poor readers (N=10)	P
Words	7.65	7.82	87
Phrases	7.18	6.41	48
Paragraphs	7.87	7.69	90
	Discrete free association test		
	Reaction time per response (100's of a sec.)		
	Good readers (N=10)	Poor readers (N=10)	P
Ave. mean time	237	220	47
Ave. median time	196	188	66

The high values obtained for P combined with the fact that three of the above differences favor the poor readers while the other two favor the good readers make it obvious that such variations as are exhibited here are almost entirely due to chance. Neither are the differences between the various types of stimuli sufficiently large to be reliable. Apparently richness and rapidity

of response is little affected either by variations in reading ability or by variations in the type of stimulus. More pertinent to the original purpose of the study, however, is the fact that in so far as the free association technique is a valid measure of richness of concepts no relationship between this attribute of the concept and reading ability has been demonstrated in this series.

The relative richness of abstract and concrete concepts. As has been pointed out, all the techniques used in this phase of the study were so constructed as to enable us to compare both abstract and concrete concepts of the two types of readers from the angle of richness. Such an analysis is contained in Table VIII. It will be noted that differences between the relationship of the two categories as well as differences within each category singly have been considered in this study.

TABLE VIII. *Logical analysis of responses on free association tests*

Continuous free association test (preliminary experiment)

	Good readers (N=14)	Poor readers (N=27)	Diff. / S.E. diff.
Concrete stimuli	8.3	7.09	1.92
Abstract stimuli	7.27	5.93	2.14
Difference	1.03	1.16	.63

Continuous free association test (main experiment)

Ave. no. associations per stimulus word

	Good readers (N=10)	Poor readers (N=10)	P
Concrete words......	8.24	8.80	
Abstract words......	7.06	6.84	
Difference......	1.18	1.96	21
Concrete phrases.....	8.50	8.00	
Abstract phrases.....	5.86	4.82	
Difference......	2.64	3.18	40
Concrete paragraphs..	8.40	8.36	
Abstract paragraphs..	7.34	7.02	
Difference......	1.06	1.34	55

Discrete free association test

Ave. reaction time per response (100's of a sec.)

	Good readers (N=10)	Poor readers (N=10)	P
Concrete words......	262	237	40
Abstract words......	213	202	62
Difference......	40	35	45

No reliable differences are indicated, either in the various single categories or in those wherein relative differences between the two categories are involved. Generic differences are no more in evidence here than they were in Series A and B.

Although not particularly pertinent to the immediate problem, it is of interest to note the reaction times of the groups as a whole on the two types of stimuli. A reorganization of the data appearing in Table VIII from this point of view yielded the results contained in Table IX.

TABLE IX. *Logical analysis of responses on free association tests*

Continuous free association test (preliminary experiment)—41 *O*s
Average number associations per stimulus word

Concrete words	Abstract words	$\frac{\text{Diff.}}{\text{S.E. diff.}}$
7.51	6.38	2.27

Continuous free association test (main experiment)—20 *O*s
Average number associations per stimulus word

Concrete		Abstract	P
	Words		
8.52		6.96	05
	Phrases		
8.26		5.32	01
	Paragraphs		
8.38		7.18	24

Discrete free association test—20 *O*s
Average reaction time per response (100's of a sec.)

Concrete words	Abstract words	P
208	250	02

While two of the above differences do not entirely fulfill the demands of significance, the three remaining differences are highly reliable and all five point in the same direction, *i.e.,* toward the greater richness of concrete concepts of both types of readers, accepting richness and rapidity of response on free association tests as a criterion of richness of concepts.

It might be argued from the fact that the 25 abstract stimuli used in the discrete free association test contained a preponderance of adjectives while the 25 concrete terms consisted mainly of nouns that the above variations are due to a difference in the grammatical rather than the logical form of the stimuli. An investigation of this point revealed, however, that the average time of reaction to adjectives in the discrete association test was almost identical with the average time of reaction to nouns, regardless of the logical form of the stimulus. *Jung* (21) also finds the time of reaction to these two forms of stimuli to be approx-

imately equal. Incidentally it argues for the reliability of the results obtained throughout this series that they are in essential agreement with those obtained by *Jung* in his extensive studies in word-association. The absolute values obtained in this study tend to be higher than those quoted by him, but the relationships of the various categories are almost identical in the two studies.

Series D: richness of concepts and reading efficiency within specific fields of knowledge. Realizing that the free association technique is not an irrefutably valid measure of richness of concepts, a further check on the results of Series C seemed desirable. An opportunity to make such a check was provided in the following study, which has as its object the determination of the relationship existing between richness of concepts in a specific field of knowledge and the ability to read material taken from that field.

Procedure. The data for the study were contained in the scores on certain of the Iowa Placement Tests (**33**). These examinations are divided into two series, a training series and an aptitude series, each comprising tests in several different academic subjects. The forms included in the training series are essentially information tests, measuring the breadth of the individual's knowledge (or, for the purposes of this study, the richness of his concepts) in various fields. The aptitude series, on the other hand, measures capacities essential to work in these fields. One section of each aptitude examination is given over to a reading test designed to measure the testee's ability to read and understand unfamiliar material taken from the field of knowledge with which the test is concerned. Thus, by determining the degree of correlation between scores on the total training examination and scores on the reading section of the related aptitude series, it was possible to state in quantitative terms the degree of relationship existing between richness of concepts and reading ability within rather definitely specified fields of knowledge.

Such an analysis was carried out in the mathematics and chemistry series of these examinations. The reliabilities of the parts utilized in this study were: chemistry training, .93; part 3 (reading test) of the chemistry aptitude, .89; mathematics training, .88; and part 4 (reading test) of the mathematics aptitude, .68.

Results. The results of this study are summarized in Table X.

TABLE X. *Correlations between richness of concepts and reading efficiency in the fields of mathematics and chemistry*

Mathematics series (N=155)

1. Training test
2. Reading test (Part 4, Aptitude Series)
3. Intelligence (Qualifying Examination)

$r12=.53\pm.04$	$r13=.65\pm.03$	$r23=.75\pm.02$	$r12.3=.08\pm.05$

Chemistry series (N=142)

1. Training test
2. Reading test (Part 3, Aptitude Series)
3. Intelligence (Qualifying Examination)

$r12.=.47\pm.04$	$r13=.63\pm.03$	$r23=.60\pm.04$	$r12.3=.15\pm.06$

While there is a high gross correlation between richness of concepts within certain fields of knowledge (in this case, mathematics and chemistry) and the ability to read material taken from these fields, the relationship is apparently mediated through a third factor, intelligence. When this variable is held constant, the net relationship between richness of concepts and reading ability becomes insignificant.

The similarity between the content measured by the Qualifying Examination and that measured by the training series of the Placement Examinations tends unduly to magnify the correlation between scores on the training test and intelligence. It thus serves to reduce the net correlation between richness of concepts and reading ability more than it would be if we had a more perfect measure of intelligence. It is questionable, though, if a correction of this defect in our measuring instruments would raise the net correlations enough to make them at all significant.

Series E: two minor studies. Two additional minor studies are of interest for their provision of additional substantiation for the results of the two preceding series.

Fertility of visual imagination as a measure of conceptual richness. Upon the assumption that fertility of visual imagination is dependent in part at least upon an adequate supply of concepts, the *Whipple* ink-blot test of imagination (**38**) was administered to the 20 *O*s of the two experimental groups. The first 11 cards of the series were used, the *O*s being instructed as follows:

"I have here a series of 11 odd-shaped ink-blots. I want you to turn them over, one at a time, and tell me all the things you see in each blot. Try them

in different positions. Of course, these blots are not really intended to represent anything in particular, but I want you to see whether your imagination will suggest pictures of things in them, just as you sometimes try to see what objects you can make out of clouds. You will have 30 seconds to look at each blot and tell me all the things you see in it. Let's try the first one to make sure that you understand what you are to do."

The following results were obtained:

TABLE XI. *Results of ink-blot test*

No. objects reported for 10 blots		
Good readers (N=10)	Poor readers (N=10)	P
31.8	34	79

In so far as the results obtained in this series are at all indicative of richness of concepts, they are in essential agreement with the results of the two preceding experiments.

Fluency of ideation as a measure of conceptual richness. The data for the second of these minor studies consisted of a series of three themes written by the members of the entering class at the University of Iowa in the fall of 1931. Assuming that fluency of ideation is indicative to a certain extent at least of the richness of concepts, the average number of ideas expressed in the three themes written by each individual was compared with the score of the same individual on part I of the Iowa Silent Reading Test.

The topics developed in these themes were " What makes life worth living ", "A book that has influenced my thinking ", and "A period of history in which I would liked to have lived ". While these subjects were not particularly well adapted to the purpose of this study, the data provided by the analysis were used only as contributory evidence. The conditions attending the writing of the compositions were fairly constant, all individuals writing on the same subject at the same time, and all work being done within a given length of time during three class periods. No preliminary warning of any of the assignments was given.

Whipple (38) suggests that " the relative number of ideas elaborated by different S's (in writing a theme) is indicated with fair approximation by the relative number of words written ". It occurred to the investigator that the number of sentences might be an even better index to the number of ideas. For purposes of comparison both indices were utilized in this analysis. The

themes of 50 persons selected at random from the group of approximately 700 who completed the assignment were examined, with the following results.

TABLE XII. *Correlation of average number of ideas expressed with scores on Part I, Iowa Silent Reading Test*

	Pt. I, I.S.R. Test
Ave. no. sentences per theme (3 themes)	.024±.095
Ave. no. words per theme (3 themes)	.053±.095

In so far as fluency of ideation is indicative of richness of concepts, little or no relationship between richness of concepts and reading ability has been demonstrated in this study.

Summary and conclusions. Accepting richness and rapidity of response on the free association test as a valid measure of conceptual richness, the performances of good and poor readers on various free association tests, both continuous and discrete, were compared. In no case was a significant difference between the two groups indicated, regardless of whether words, phrases, or paragraphs were utilized as stimulus materials.

Neither were reliable differences indicated by further experimentation. Using scores on the training examinations of the Iowa Placement Tests as measures of conceptual richness and scores on the reading tests included in the related aptitude series as indicators of reading ability, the correlations between these two functions in two rather specific fields of knowledge were determined. While the gross correlations were significant (chemistry, .47; mathematics, .53), the relationship was found to be mediated through a third factor, intelligence. When this variable was held constant the net correlations dropped to .15 and .08, respectively, indicating a comparative lack of relationship between richness of concepts and reading ability within at least two particular fields of knowledge.

Two additional minor studies also contributed to this phase of the investigation. In the first, *Whipple's* ink-blot test of visual imagination was administered to the 20 *O*s of the two regular experimental groups, upon the assumption that fertility of visual imagination is dependent in a certain sense upon the richness of concepts. The second was based on an analysis of the themes

in different positions. Of course, these blots are not really intended to represent anything in particular, but I want you to see whether your imagination will suggest pictures of things in them, just as you sometimes try to see what objects you can make out of clouds. You will have 30 seconds to look at each blot and tell me all the things you see in it. Let's try the first one to make sure that you understand what you are to do."

The following results were obtained:

TABLE XI. *Results of ink-blot test*

	No. objects reported for 10 blots	
Good readers (N=10)	Poor readers (N=10)	P
31.8	34	79

In so far as the results obtained in this series are at all indicative of richness of concepts, they are in essential agreement with the results of the two preceding experiments.

Fluency of ideation as a measure of conceptual richness. The data for the second of these minor studies consisted of a series of three themes written by the members of the entering class at the University of Iowa in the fall of 1931. Assuming that fluency of ideation is indicative to a certain extent at least of the richness of concepts, the average number of ideas expressed in the three themes written by each individual was compared with the score of the same individual on part I of the Iowa Silent Reading Test.

The topics developed in these themes were " What makes life worth living ", "A book that has influenced my thinking ", and "A period of history in which I would liked to have lived ". While these subjects were not particularly well adapted to the purpose of this study, the data provided by the analysis were used only as contributory evidence. The conditions attending the writing of the compositions were fairly constant, all individuals writing on the same subject at the same time, and all work being done within a given length of time during three class periods. No preliminary warning of any of the assignments was given.

Whipple (38) suggests that " the relative number of ideas elaborated by different S's (in writing a theme) is indicated with fair approximation by the relative number of words written ". It occurred to the investigator that the number of sentences might be an even better index to the number of ideas. For purposes of comparison both indices were utilized in this analysis. The

themes of 50 persons selected at random from the group of approximately 700 who completed the assignment were examined, with the following results.

Table XII. *Correlation of average number of ideas expressed with scores on Part I, Iowa Silent Reading Test*

	Pt. I, I.S.R. Test
Ave. no. sentences per theme (3 themes)	.024±.095
Ave. no. words per theme (3 themes)	.053±.095

In so far as fluency of ideation is indicative of richness of concepts, little or no relationship between richness of concepts and reading ability has been demonstrated in this study.

Summary and conclusions. Accepting richness and rapidity of response on the free association test as a valid measure of conceptual richness, the performances of good and poor readers on various free association tests, both continuous and discrete, were compared. In no case was a significant difference between the two groups indicated, regardless of whether words, phrases, or paragraphs were utilized as stimulus materials.

Neither were reliable differences indicated by further experimentation. Using scores on the training examinations of the Iowa Placement Tests as measures of conceptual richness and scores on the reading tests included in the related aptitude series as indicators of reading ability, the correlations between these two functions in two rather specific fields of knowledge were determined. While the gross correlations were significant (chemistry, .47; mathematics, .53), the relationship was found to be mediated through a third factor, intelligence. When this variable was held constant the net correlations dropped to .15 and .08, respectively, indicating a comparative lack of relationship between richness of concepts and reading ability within at least two particular fields of knowledge.

Two additional minor studies also contributed to this phase of the investigation. In the first, *Whipple's* ink-blot test of visual imagination was administered to the 20 *O*s of the two regular experimental groups, upon the assumption that fertility of visual imagination is dependent in a certain sense upon the richness of concepts. The second was based on an analysis of the themes

of 50 individuals, the average number of ideas developed in the three themes written by each person being accepted as a measure of the richness of his concepts. In the former study no reliable difference between the average number of objects reported by the two groups was indicated; in the latter the correlation between the number of ideas expressed in the three themes and scores on the Iowa Silent Reading Test was found to be insignificant.

In so far as these techniques are valid measures of the richness of concepts, none of them have brought to light any true differences between good and poor readers in this respect.

The association tests used in this phase of the study were also constructed in such a way as to make possible a further study of generic differentiæ, this time from the point of view of the differential richness of abstract and concrete concepts. For the third time, however, no such variations were revealed. Striking differences were noted between the reactions of the two groups as a whole to the two types of stimuli, but no such variations from one group of readers to the other were discernible.

Organization of the concepts of good and poor readers. While the exact nature of the attribute of organization can not be so clearly and definitely specified as could the nature of richness, it is assumed that this characteristic of the concept may be identified with the availability of concept-meanings for immediate use, or with the closeness of association between the concept-symbol and the meanings which it symbolizes. Such a definition would appear to be warranted by the meaning usually attributed to the term. We commonly regard organization as that principle whereby a collection of parts is so ordered as to possess added utility. Similarly, conceptual organization is reflected in a more ready availability of specific items for use upon demand.

Series F: the controlled association test as a measure of organization. In the light of this interpretation of the term, the controlled association technique was selected as best adapted to the demands of the investigation, the degree of organization being assumed to vary inversely with the length of the reaction-time.

Preliminary experimentation. In order to familiarize himself with the controlled association technique, as well as to determine

the validity of the instrument for use in this situation, the investigator conducted a short preliminary experiment, using three controlled association tests, all adapted from the *Woodworth-Wells* series of association tests (40). The two groups utilized in the preliminary phase of Series C served as *O*s. The poor group was supplemented with some 17 individuals; however, the original homogeneity of the group was little affected by this change.

The 20 words of each test were typed on a long, narrow sheet of paper. This was placed before *O* face down, while he was instructed as specified in the *Woodworth-Wells* monograph (40). After making sure that *O* had the instructions clearly in mind, the stimuli were exposed and the total time required to respond to the entire 20 words was measured with a stop-watch. Omitted and incorrect responses were also recorded.

The results of this experiment are recorded in Table XIII.

TABLE XIII. *Results of preliminary experiment with controlled association tests*

Ave. reaction time per response (in sec.)

	Good readers (N=14)	Poor readers (N=44)	$\frac{\text{Diff.}}{\text{S.E. diff.}}$
Part-whole test...	1.57	2.32	5.21
Subject-verb test..	1.35	1.76	3.08
Analogies test....	3.08	4.50	6.19

In striking contrast to the insignificant differences revealed in the investigation on richness of concepts are the highly reliable differences indicated for all three controlled association tests. In so far as these techniques are valid measures of the attribute of organization, it is obvious that the concepts of the good readers were much better organized than those of the poor readers.

The objection may be advanced here that in as much as tests such as these are frequently used as measures of intelligence and the two groups of *O*s used in this study varied quite widely in intelligence, the above differences may have been caused by variations in intelligence rather than by variations in reading ability. An answer to this criticism is provided in the following investigation.

Procedure. The 20 *O*s of the regular experimental groups

were utilized in this study. The part-whole and analogies tests used in the preliminary experiment, both lengthened to 30 words, again served as experimental techniques, while the subject-verb test of the former investigation was replaced by an opposites test of 30 items adapted from *Means'* Hard Opposites Series (23). The tachistoscope and counter set-up described in Series C was used for the measurement of reaction-times. The procedure was essentially the same as that followed with the discrete free association test, while the instructions were adapted with minor changes from those used in the preliminary experiment.

Results. The results of this study are summarized in Table XIV. Here again the occasional occurrence of excessively long reaction-times made it necessary to compute both median and mean reaction times for each individual.

TABLE XIV. *Results obtained with controlled association tests*

	Ave. median reaction-time per response (100's of a sec.)		
	Good readers (N=10)	Poor readers (N=10)	P
Part-whole test....	163	192	03
Opposites test......	195	223	07
Analogies test......	276	313	07
	Ave. mean reaction-time per response (100's of a sec.)		
	Good readers (N=10)	Poor readers (N=10)	P
Part-whole test....	223	249	27
Opposites test......	296	381	02
Analogies test......	341	380	18

While only two of the above values of P meet the rigid requirements of reliability, the weight of evidence as a whole points toward the more effective organization of the concepts of good readers as compared to those of less efficient readers. The tendency is more pronounced in the case of median than mean reaction times. If the investigator were to choose between the two sets of averages, however, he would unhesitatingly select the former as the more truly representative measures.

It should be noted that the differences revealed in this study can not be accounted for on the basis of intellectual variations among the *O*s. It is true that the differences indicated here are not so highly reliable as those of the preliminary investigation. They possess a sufficient degree of significance, however, to war-

rant the conclusion that some functional variant of reading ability other than general intelligence is being measured by this technique. Upon the basis of the assumption outlined at the beginning of this section, the writer is inclined to believe that this variant is conceptual organization.

Series G: the completion test as a measure of organization. While the results of Series F were conclusive enough, a desire to determine whether a different approach to the problem of organization would yield the same results led to a second investigation in this phase of the study. The completion test included in the 1929 edition of the American Council of Education Psychological Examination, wherein the testee is required to complete 40 definitive statements with words indicating the concepts being defined, served as the means of measurement. The assumption was that an adequate performance on this test is dependent upon the proximity of association between various concept-symbols and their respective meanings. This latter function has already been identified with the attribute of conceptual organization. The concepts involved in the test are sufficiently familiar so that the instrument actually measures the proximity of already existing associations between concept-symbols and meanings and not merely the number of such associations, as does the common vocabulary test.

Procedure. Here again the 20 individuals of the experimental groups served as *O*s. The standard instructions were followed in the administration of the test, except for the fact that the usual 10-minute time-limit was increased to 15 minutes, in order to give all the *O*s time to get over the entire 40 items at least once. While this change may have lowered the reported reliability of the test (.834), it is felt that the resulting increase in validity more than compensated for any such loss.

Results. Table XV summarizes the results of this study. The averages for both the number of items attempted and the number of items correctly answered are recorded. This was done in order to meet the criticism that in the latter measure we are confusing accuracy and organization and therefore a purer measure of organization would be the gross number of items tried.

Both sets of findings point, however, in the same direction and are entirely consistent with those obtained in Series F. The fact that the two techniques approach the problem from quite different angles only serves to make the consistency more striking.

TABLE XV. *Results of completion test*

	Good readers (N=10)	Poor readers (N=10)	P
Ave. no. items correct...	25.8	16.5	00
Ave. no. items attempted	28.8	20.8	00

Summary and conclusions. Utilizing rapidity of response on various controlled association tests and scores on a definitions test of the completion type as measures of organization, the concepts of a group of good and a group of poor readers were compared. The results of both experiments point unreservedly in the same direction, *i.e.,* toward the more effective organization of good readers' concepts as compared with those of less efficient readers. This conclusion is contingent, of course, upon the validity of these techniques as measures of conceptual organization.

Clarity of the concepts of good and poor readers. Series H: assurance reports as measures of clarity. Due to the fact that the conceptual attribute of clarity is the least objective of all the characteristics postulated of the concept, some difficulty was encountered in securing a satisfactory instrument for measuring it. *Kakise's* statement that " clearness or unclearness of meaning, abstracted from content-feeling or reproduction, may be reduced (introspectively) to mere feelings of certainty and uncertainty " (22) appeared, however, to provide experimental justification for the measurement of clarity in terms of the more easily observable experience of certainty or uncertainty. Consequently, an approach to the problem from this point of view was settled upon.

The measuring instrument. After a consideration of various vocabulary tests, the first 80 items of Form N of the College Aptitude Test used by the Association of Minnesota Colleges, which consists of a series of vocabulary tests, was selected as the basic instrument for use in this experiment. This particular test was chosen because the level of difficulty was judged to be such

as would obtain the most valid results with the *O*s utilized here. The estimated reliability of this form was also higher (approximately .80) than that found for any other vocabulary test of the same length.

Using this test as a basic technique, then, a supplementary scale for indicating the degree of certainty accompanying the reaction to each of the 80 items attempted was devised. Three points on this scale besides complete uncertainty, in which case *O* indicated no response whatever, were distinguished. These, as described by *Whipple* (38) in his discussion of methods of grading the assurance of reports in the *Aussage Test,* were: "(1) hesitancy; (2) positive statement or assurance of ordinary degree; and (3) attestation or attested assurance, *i.e.,* the highest degree of assurance".

Procedure. The test was administered to the 20 *O*s of the experimental groups. They were instructed as follows:

"The object of this test is, first, to find out how many words you know, and secondly, to find out how well you know them. After each word in capital letters there are five answers, only one of which defines it. Pick out the right answer and draw a line under it for reference. Then put the number of that answer in the parentheses at the end of the line, as shown in the samples. After you have done that, write 1, 2, or 3 outside the parentheses to show how certain you are of the correctness of your answer. The following scale indicates the degrees of certainty indicated by these numbers:

1. I believe that this is the correct meaning of the word, but I am not at all certain.
2. I am fairly certain that this is the correct meaning, but I am not altogether positive.
3. I am absolutely positive that this is the correct meaning.

In other words, 1 indicates that your response is only a little better than a guess; 2 indicates that, while you are fairly certain of the correctness of your answer, you are not altogether positive; and 3 indicates that you are as certain of the correctness of your answer as it is ever possible for anyone to be. Of course, if you do not know the meaning of the word at all, record nothing either inside or outside the parentheses.

You are to do two things, then, for each of the words in this test. First, write in the parenthesis the number of the answer that best defines the word; and second, write outside of the parentheses that number (1, 2, or 3) which best indicates the degree of assurance accompanying your response. You will have 15 minutes to go through the test. Let's try a few samples to make sure that you understand what you are to do."

After presenting three or four samples to make sure that *O* understood the procedure to be followed, as well as the significance of the three steps in the assurance scale, he was given the signal,

" Go ". The stop-watch was started simultaneously and time was called at the end of 15 minutes.

Results. An average " assurance score " was computed for each individual by totaling the " assurance responses " and dividing by the number of items attempted. The means of the ten such averages obtained for each group are indicated in Table XVI. The second item in the table is included to show that in this test, as well as in part II of the I.S.R. test, the two groups of *O*s are demonstrated to vary significantly in vocabulary skill.

TABLE XVI. *Average assurance and accuracy scores on the Minnesota Vocabulary Test*

	Good readers (N=10)	Poor readers (N=10)	P
Ave. degree of certainty (on a scale of 3)	2.60	2.38	01
Ave. no. items correctly answered........	57.5	36.1	00

While the gross difference between the average " assurance scores " of the two groups is not so striking, it is unquestionably significant. Statistically stated, the chances are 99 in 100 that the variation is a true one.

Summary and conclusion. In this investigation the concepts of a group of good and a group of poor readers were compared from the standpoint of clarity. Accepting the average of a series of judgments relative to the degree of assurance accompanying responses to 80 items in a vocabulary test as a measure of this attribute, the concepts of the good readers were found to possess a significantly greater degree of clarity than those of the poor readers. While no additional evidence was available on this point, the technique was sufficiently pertinent and the results clear-cut enough to warrant the generalization that conceptual clarity is an important factor in reading ability.

The accuracy of the concepts of good and poor readers. The significance of the attribute of accuracy is sufficiently obvious to require no further elaboration. For all practical purposes, the degree of accuracy possessed by a given concept is determined by the extent to which the meanings included within it agree with the meanings given for that concept in a standard dictionary. Theoretically, varying degrees of accuracy all the way from complete and sheer wrongness to perfect correctness may be distinguished;

it is difficult, however, to locate any very definite intervals between these two extremes.

Series I: the ability to discriminate between closely related concepts as a measure of accuracy. In casting about for means of measuring conceptual accuracy the investigator was struck by the applicability of a familiar intelligence test technique, which involves the discrimination of differences between closely related concepts. Whatever else may be measured by this test, an adequate performance on it certainly presupposes a supply of concepts possessing accurate, clear-cut meanings.

Preliminary experimentation. The usual method of administering this test is to ask *O,* without further preliminaries, to point out the differences between given pairs of terms " in your own words ". Although skeptical of the reliability of this procedure, especially when used as an isolated test, a preliminary experiment was undertaken with the technique in this form.

A series of 30 pairs of terms was constructed. While a few of them were adapted from standard tests, the great majority were selected from a standard dictionary, from word books, and from books of synonyms. In every case, an attempt was made to select pairs of concepts differing in only one respect. Too obvious differences were avoided; on the other hand, pairs of terms exemplifying exceptionally fine shades of meaning, or differences such as would be known to only a small number of persons, were also rejected. Only items involving distinctions made by accepted authorities were used. All cases wherein there was a difference of opinion as to the validity of the distinction were automatically rejected.

This series of 30 items was presented to each of the 20 *O*s of the two experimental groups with the instruction to " distinguish between each of these pairs of terms in your own words ". No time limit was set on the test. A stenographic report of *O*'s responses was taken down by *E*.

An attempt at an analysis of the reports thus obtained only served to confirm the investigator's suspicions. In spite of all attempts to secure concise, intelligible answers, most of the reports were so ambiguous as to make any analysis of them practically worthless. For this reason, while the general form of the

technique was retained, a more objective method of administering the test was devised for use in the main experiment.

In this revised form a multiple choice technique was utilized. The test was made up, as before, of 30 items. Here, however, four alternative responses were devised for each item and *O* was required to select that statement that best expressed the difference between the terms. While certain advantages of the former method of administration were sacrificed in this change, it was felt that such losses were more than offset by the gain in objectivity attendant upon the utilization of this method.

The four alternative responses were made up of three statements indicating possible differences between the two items, while the fourth read, " no difference ". Accordingly, five items of the test were pairs of concepts possessing identical meanings. The remaining 25 consisted of those items which had been demonstrated roughly by the preliminary experiment to be most discriminative and enough more new items to make up the required total.

Procedure. *O* was first requested to read through the instructions found at the beginning of the test. After making sure that he had the procedure clearly in mind *E* gave the signal " Go " and the stop-watch was started. Time was called at the end of 15 minutes.

Results. The results of this experiment are contained in Table XVII.

TABLE XVII. *Analysis of scores on " Differences Test "*

Ave. no. items answered correctly		
Good readers (N=10)	Poor readers (N=10)	P
19.2	17.5	24

No significant difference between the average scores of the two groups on this test is indicated. The inability of the technique to discriminate more clearly between the two types of readers is probably due in part, however, to the low reliability of the test (.42 by the chance-halves method). At any rate, in the light of a knowledge of the results obtained with other techniques in this phase of the study, the writer feels called upon to point out that a rather distinct tendency is indicated by the results of this study.

He is also convinced that this technique deserves further consideration as a measure of conceptual accuracy. Such a test of acceptable reliability would be of use in numerous situations.

Series J: accuracy of response on vocabulary and completion tests. Procedure. A second approach to the measurement of the accuracy of concepts was afforded by a further analysis of three sets of data already at hand. Two consisted of scores made by the two groups of *O*s, first, on part II of the Iowa Silent Reading Test, and secondly, on that section of the Minnesota College Aptitude Examination which was utilized in Series H for the measurement of conceptual clarity. The performances of the 20 *O*s on the completion test used in Series G were also capable of analysis from this point of view.

Instead of accepting, as indicators of conceptual accuracy, the raw scores on these tests, wherein the two groups have already been shown to vary in a highly significant manner, the percentage of the number of items attempted that were correct was computed as an " accuracy quotient ". While the validity of this procedure may be questioned, it seemed to the investigator to be justified on the grounds that it takes cognizance of both the number of items attempted and the number correct, while the usual method of computing scores takes cognizance only of the gross number of items that are correct. For example, the latter method fails to indicate whether a low score is the result of a large but inaccurate vocabulary or a small but accurate vocabulary. The effect of the method utilized in this analysis is to place all vocabularies on a par as far as size is concerned and then measure the relative accuracy of each. In other words, it isolates a single characteristic of the concept, accuracy, and measures it apart from any others which may be involved in the test.

Results. Such an analysis of the scores of the 20 *O*s on the three tests mentioned above is summarized in Table XVIII.

TABLE XVIII. *Accuracy of responses on vocabulary and completion tests*

	Pct. of items attempted that were correct		
	Good readers (N=10)	Poor readers (N=10)	P
I.S.R., Pt. II........	81.1	77.0	10
Minnesota Test......	80.1	69.4	00
Completion Test.....	89.5	79.6	01

While the first of the above differences is not of sufficient magnitude to meet the demands of reliability, the last two are highly significant and all point toward the greater accuracy of the concepts of the good readers. These results, taken in conjunction with those obtained in Series I, provide in themselves a strong case for the generalization that efficiency in reading is dependent to a certain extent upon the accuracy as well as the clarity and organization of concepts. Even further substantiation of this conclusion was forthcoming.

It is of some interest to note the questionable significance of the difference between the average scores of the two groups on part II of the Iowa Silent Reading Test when computed in this manner as compared to the highly reliable difference obtained when the scores were computed in the ordinary manner. As far as this group of individuals was concerned, gross scores on this test were apparently more indicative of differences in the size than of the accuracy of their respective vocabularies.

Series K: the selection of correct meanings as a measure of accuracy. One last bit of evidence, which is in essential agreement with the results of the two preceding series, was afforded by an analysis of the performances of the two groups of readers on section 4 of the English Training Test of the Iowa Placement Series. In this examination, the testee is required to select from three alternative statements that one which most correctly illustrates the specific meaning of a given word. Except for the fact that specific meanings of single concepts rather than distinctions between pairs of concepts are involved here, the test is quite similar to the instrument described in Series I. For this reason, the following analysis is interesting as an indicator of the results that would have probably been obtained in the previous experiment had we possessed a more reliable form of the test used therein.[7]

Results. Table XIX summarizes the results of an analysis of the scores of the 20 *O*s of the two experimental groups on this examination. Since this technique was originally devised, as was the one utilized in Series I, to measure accuracy alone, it was

[7] The reported reliability of this part of the English Training Test is .886.

necessary to consider only the gross number of items correctly answered.

TABLE XIX. *Ave. scores on section 4 of English Training Test*

Average number items correctly answered		
Good readers (N=10)	Poor readers (N=10)	P
39.8	27.4	00

Since scores on both this test and the Iowa Silent Reading Test were available for the entire freshman class of 1931, the two sets of scores were correlated to find out whether or not such an approach would bear out the results indicated in Table XIX. Such a treatment of the scores of 100 individuals selected at random from the larger group revealed a correlation of .57 ± .04. This figure was reduced to .26 ± .06 when intelligence, as measured by the Qualifying Examination, was partialed out. While this is not a high relationship, it is significant, thus bearing out the results indicated in the above table. Both sets of data agree, of course, with Series I and J in demonstrating a real relationship between reading ability and conceptual accuracy.

Summary and conclusions. A study of the accuracy of good and poor readers' concepts was taken up from several different points of view in this phase of the investigation. After preliminary experimentation designed to reveal the most adequate method of procedure, a multiple choice form of the familiar test technique wherein *O* is required to discriminate between closely related concepts was constructed and administered to the 20 *O*s of the experimental groups. While a significant difference between the two types of readers was not indicated, a tendency which was adequately substantiated by further experimentation was revealed.

A second attack on this problem was provided by an accuracy analysis of the performances of the two groups of readers on two vocabulary tests and the completion test used in Series G. One of the differences revealed by the analysis failed to meet the demands of significance; the other two, however, were unquestionably reliable and all three pointed toward the inherence of a greater degree of accuracy in the concepts of the good readers.

The weight of evidence in the third study pointed in the same direction as the two preceding. Accepting the number of items correctly answered on part 4 of the English Training Test of the Iowa Placement Series as a measure of conceptual accuracy, the average score of the 10 good readers in this test was found to be reliably higher than that of the 10 poor readers. The correlation of the scores of 100 individuals on this test against the scores of the same individuals on part I of the Iowa Silent Reading Test, with intelligence partialed out, was also found to be significant ($.26 \pm .06$).

In general, these investigations provide ample evidence for the conclusion that the concepts of good readers are more accurate than those of less efficient readers. Here again, of course, these conclusions are warranted only to the extent to which the techniques utilized in this phase of the study are valid measures of conceptual accuracy. In most cases, however, the pertinence of the instrument to the problem at hand was obvious.

VI. General summary and conclusions. In an effort to determine which characteristics of the concept are most essentially involved in reading ability, the concepts of ten good readers were compared with those of ten poor readers. A preliminary analysis of the concept indicated that characteristics of both a qualitative and quantitative nature should be considered, with the latter category including the qualities of richness, organization, clarity, and accuracy. Therefore, techniques designed to test the concepts of the above mentioned readers from these various points of view were utilized.

The investigations of Series A and B were aimed at a study of possible generic differences between the concepts of the two types of readers, the former from the standpoint of the mental content of concepts, the latter from the angle of a comparative study of the types of responses exhibited in a free association test. Three or four possible points of variation were revealed by the introspective technique utilized in Series A. The weight of evidence as a whole was, however, clearly in the direction of structural similarity rather than dissimilarity between the conceptual con-

sciousnesses of the two groups of readers. The results obtained in Series B pointed in the same direction. In none of the six categories into which the responses of the *O*s were classified were any reliable differences between the two groups of readers demonstrated. A third approach to the problem of generic differentiæ was afforded by a comparison of the abstract and concrete concepts of the two classes of readers, from the two points of view assumed in Series A and B. The results of this comparison only served to provide further substantiation for the conclusions of the two previous studies. In so far as the techniques utilized in this phase of the study are valid, variations in reading ability, in the extremes at least, were not shown to be significantly affected by conceptual variations of a generic nature.

Incidentally it was of interest to note in connection with the third approach mentioned above that when we disregarded the good reader–poor reader dichotomy and analyzed the data from the standpoint of responses to abstract and concrete stimuli, striking differences were apparent, particularly in the results of Series B. It was decided, however, that such variations were probably predetermined by external rather than by internal factors, *i.e.*, by the manner in which the test instrument was constructed rather than by any factors inherent within the mental makeup of the *O*s.

Series C, D, and E were devoted to investigations of the influence of the conceptual attribute of richness upon reading ability. Utilizing free association tests, both continuous and discrete, as measures of richness, Series C brought to light no differences of any significance whatever between the two types of readers. In a preliminary experiment, wherein wide differences in intelligence as well as in reading ability occurred, a tendency for the good readers to surpass the poor readers in the average number of responses made on a continuous free association test was noted. When the factor of intelligence was held constant, however, as it was in the main experiment, variations of a purely chance order were obtained. In Series D the problem was approached from the angle of the relationship existing between the richness of concepts in a specific field of knowledge and the ability to read

unfamiliar material taken from that field. The data were provided by scores on certain of the Iowa Placement Tests. A high gross correlation was found to exist between the two factors. Further study of the data revealed, however, that the relationship was mediated through a third factor, general intelligence, and when this variable was held constant, the real relationship between richness of concepts and reading ability within the specific fields here considered was of a negligible order. In addition to supporting the conclusion of Series C, this finding emphasizes the necessity of ruling intelligence out of the picture when making any analysis of the factors involved in reading ability. In all probability many of the psychological mechanisms which have been postulated as fundamental determiners of reading efficiency function as such, not in and of themselves, but rather through the medium of general intelligence. Series E consisted of two minor studies which, taken by themselves, would be of questionable value, but which were of interest in their provision of evidence which served further to clinch the conclusions of the two preceding studies. In the first, wherein scores on the Whipple Ink-Blot Test were utilized as measures of conceptual richness, the difference between the average scores of the two groups of readers was of a strictly chance order. In the second the numbers of ideas developed by a random assortment of 50 individuals on three themes were correlated against the scores made by these same individuals on a standard reading test. Correlations of an insignificant order were obtained. Every line of evidence utilized in this phase of the investigation pointed in the same direction, *i.e.*, toward the questionable existence of a true relationship between conceptual richness and reading ability. Certainly no one of the techniques here utilized has provided a whit of evidence that reading efficiency is affected in any significant manner by the richness of concepts.

A further study of generic differences, this time from the point of view of the relative richness of abstract and concrete concepts, was also made possible by the peculiar manner in which the free association tests of Series C were constructed. While no reliable

variations from one type of reader to the other were evidenced, significant differences from one type of stimulus to the other (disregarding again the distinction between the two groups of *Os*) were indicated, in the direction of the greater richness of concrete concepts. The possibility that this variation was attributable to grammatical rather than logical differences in the form of the stimuli was adequately ruled out. It was noted that this finding, as were certain others obtained through the use of the free association test, was in essential accord with those obtained by *Jung* in his studies in word association.

In Series F and G attempts were made to appraise the possibility that reading efficiency might be conditioned in a measure upon the degree of organization inhering within the concepts of the reader. The interpretation placed upon the term organization appeared to point in the direction of the controlled association technique as the most valid instrument available for measuring this characteristic of the concept, the degree of organization being assumed to vary inversely with the length of the reaction time. The results of Series F, wherein the controlled association technique was utilized, indicated that conceptual organization plays a significant rôle in reading ability. Although not all the differences obtained were found to be significant, a majority of them were within the limits of reliability and all pointed toward the superior organization of the concepts of good readers. In a preliminary experiment, wherein the factor of general intelligence was not controlled, the differences were of a higher order of reliability than those obtained in the main experiment. The results of the latter investigation were sufficiently significant, however, to indicate that some functional variant of reading ability other than general intelligence was being measured. For reasons already indicated the writer is inclined to believe that conceptual organization constitutes this variant. The results of Series G also pointed toward the more adequate organization of the concepts of good readers. Using a definitions test of the completion type as a measure of the attribute of organization, the average performance of the good readers was shown to be significantly

superior to that of the poor readers, whether the average number of items correct or the average number of items attempted be used as an index of the level of efficiency. The warranted conclusion of this phase of the investigation would seem to be that a real functional relationship exists between reading ability and conceptual organization. This conclusion is contingent, of course, upon the validity of the techniques used as measures of the latter characteristic.

The clarity of the concepts of good and poor readers was the point of emphasis around which Series H revolved. Utilizing judgments relative to the degree of assurance accompanying responses to items in a vocabulary test as measures of clarity, the concepts of good readers were found to possess a significantly greater degree of clarity than those of poor readers. While further evidence on this point was not available, the results of this series were clear-cut enough to make it appear that reading efficiency is dependent in some measure upon the degree of clarity inhering within the concepts which the individual brings with him to his reading. Incidentally, the writer feels that the problem indicated in this series contains possibilities for further research. For instance, the question of the relationship between the clarity of a person's concepts and the accuracy of such concepts would provide an interesting problem for study, as would also an investigation of the factors making for variations in clarity.

Series I, J, and K were designed to determine the bearing of conceptual accuracy upon reading ability. The three approaches were unanimous in revealing the dependence of reading efficiency upon this characteristic of the concept. In the first, which involved the adaptation of a familiar intelligence-test technique to the requirements of this particular study, the difference between the average scores of the two types of readers was not significant. It pointed, however, toward a more adequate performance on the part of the better readers. The inability of the technique to discriminate more clearly between the two types of readers was probably due in part to the low reliability of the instrument. This impression received a degree of confirmation from the results of

Series K, wherein a test technique very similar to that used in Series I, but of adequate reliability, was utilized. Here unquestionably significant differences in favor of the better readers were revealed, not only by an analysis of the performances of the 20 readers composing the regular experimental groups, but also by correlating the scores made on the test by 100 randomly selected individuals against their reading scores. The results of Series J were based on an analysis of the scores of the two groups of readers on three vocabulary tests. Here, instead of raw scores as measures of accuracy, the percentage of the number of items attempted that were correct was computed as an "accuracy quotient". While one of the differences thus obtained was not of sufficient magnitude to meet the demands of reliability, the other two were highly reliable and all pointed toward the greater accuracy of the concepts of the good readers. These lines of evidence would appear to provide ample support for the conclusion that reading efficiency is conditioned to a certain extent upon the accuracy of concepts. The question of the validity of the measures utilized arises again. The pertinence of the instruments to the problem appears, however, to be fairly obvious.

By way of summarizing the study as a whole, the concept appears to play a rôle in reading ability from the point of view of organization, clarity, and accuracy. But the investigation has not demonstrated that generic variations or differences in the richness of concepts are of any particular importance as co-variants of reading ability.

VII. Clinical implications. Having arrived at these conclusions, the further question arises as to what practical implications they possess for remedial procedures in reading. Assuming that further experimentation along these lines would bear out the results obtained in this investigation, what do they mean from the standpoint of dealing with reading difficulties? Of course, the study stands or falls on its own merits, quite apart from any clinical implications it may possess. The writer feels, however, that there are a few points at which the conclusions obtained intersect with diagnostic and remedial procedures.

In the first place, it is felt that the investigation has demonstrated the inadequacy of the ordinary vocabulary test as an indicator of a simple variant affecting reading. The statement has frequently been made in the literature that vocabulary is a determining factor in reading efficiency. What aspect of vocabulary, the writer would like to inquire. The results of the present study appear to have demonstrated that the ability to use a vocabulary is not the simple capacity that we have assumed and that when we speak of the use of a vocabulary as a conditioning factor in reading ability we must specify whether we mean richness, accuracy, organization, clarity or type of vocabulary. *Pressey* (27) has hinted that vocabulary equipment may not be a simple function and this investigation appears to provide a certain amount of experimental substantiation for such a statement.

If the ordinary vocabulary test is not an adequate diagnostic instrument for use in the reading clinic, we must devise tests that will meet such demands. While the instruments used in this study have served very well, the investigator feels that additional experimentation is needed, not only to " iron out " weaknesses in these particular devices, but also to construct new ones of a more efficient nature. He is inclined to believe that his analysis of the various aspects of vocabulary skill is essentially valid. Working along these lines, the next step is to provide more efficient ways and means of measuring these functions.

Apparently it will not be necessary to consider all the qualities of the concept measured in this study. It would seem, tentatively, that only those characteristics that we have dealt with under the captions of accuracy, organization, and clarity will need to be taken into consideration, the influence of the type and richness of an individual's concepts upon his reading efficiency being so slight as to preclude them as conditioning factors. This statement, of course, awaits further confirmation. But if such experimentation should bear out the results of this study, there is no good reason why we should bother about generic differences (at least, of the type we have considered in this study) or about quantitative richness in diagnosing vocabulary shortcomings. Accuracy, organ-

ization, and clarity are the differentiæ that should primarily concern us.

The second finding of major importance for clinical work in reading is contained in the fact that the study has established no particular relationship between reading ability and the richness of that person's concepts. This finding is so diametrically opposed to the generally accepted notion that the richness of a person's vocabulary is one of the major determiners of his ability to understand language material of any kind that one is exceedingly hesitant about accepting it. But it is not inconceivable to the writer that further experimentation will confirm these results. It may be that such factors as the accuracy, clarity, and organization of concepts are actually more important for reading efficiency than sheer amount. The present study constitutes the first effort to measure this attribute of vocabulary apart from these other factors and it may very well be that our concession of a position of primary importance to the quality of richness has been due to a failure to distinguish clearly between richness and these other qualities of the concept. At least it will be valuable to undertake a further study of this point, this time from the standpoint of the more adequate analysis afforded by our systematic description of the attributive characteristics of the concept.

The third point which the writer wishes to stress in this connection concerns the rôle played by general intelligence in reading ability. In clinical work it should be one of the first items to be considered. For not only have statistical investigations shown a high correlation to exist between intelligence and reading ability, but remedial work in the clinic (**41**) has also demonstrated that individuals of low intelligence, unless highly motivated, derive little benefit from training in reading. The writer wishes to point out, however, that a clear distinction should be made between general intelligence and vocabulary skill in diagnostic work. The relationship between these two functions is so close that it is very easy to confuse them. Unless they are quite clearly distinguished the real cause of the difficulty may not be discerned and consequently can not be adequately dealt with. This point is also

important in carrying on research directed toward the analysis of factors conditioning reading difficulties. It can not be doubted that many of the capacities which have been pointed out as primary determiners of reading efficiency actually function through the medium of general intelligence and if this factor were to be held constant they would be found to bear only a minor relationship to reading efficiency. Of course, it is desirable to reduce the altogether too general function of general intelligence to lower terms whenever possible; but a recognition of the relationship between the capacity being considered and native intelligence sometimes means the difference between a futile waste of time in attempting to train the capacity and a wise economy of time in according it only passing consideration.

This brings up the question of the susceptibility of the various conceptual attributes considered in this dissertation to training. The writer feels that this question can be adequately answered only upon the basis of additional experimentation which is beyond the scope of the present study. The quantitative richness of concepts could be increased, of course, without any undue expenditure of time or effort by adding to the experiential background of the individual. But beyond a certain level, which appears to be reached more or less simultaneously with the achievement of a definite level of intellectual development, richness seems to exert a minor effect upon reading efficiency. Apparently it would be much more desirable to provide training aimed at increasing the accuracy or clarity of a person's vocabulary, or at providing him with a more adequately organized system of concepts. How to carry on training designed to improve these functions or the possibilities of obtaining improvement, were we to attempt such training, are, however, questions that can not be answered at present.

Bibliography

1. Aveling, F. The consciousness of the universal. London, Macmillan, 1912.
2. Bagley, W. C. The apperception of the spoken sentence, *Amer. J. Psychol.*, 12, 1900, 80–134.
3. Beall, R. The vocabulary of certain primary reading tests as a factor limiting their validity, Thesis (Univ. of Iowa), 1924.
4. Binet, A. L'etude experimentelle de l'intelligence. Paris, Reinwold, 1903.
5. Bird, G. E. An experiment in focalization, *School and Soc.*, 8, 1918, 569–70.

6. Bleuler, E. Upon the significance of association experiments, Studies in Word Association (Tr. by Eder). London, Heinman, 1918. Pp. 1–7.
7. Bolton, T. L. Meaning as adjustment, *Psychol. Rev.*, 15, 1908, 169–172.
8. Condit, M. The effect of familiarity of subject matter upon speed and comprehension in silent reading, Thesis (Univ. of Iowa), 1919.
9. Dewey, J. C. A case study of reading comprehension of difficulties in American history, Thesis (Univ. of Iowa), 1931.
10. Drever, J. Meaning as affective, Ninth International Congress of Psychology, Proc. and Papers, Psychol. Rev. Co., Princeton, N. J., 1930. Pp. 152–154.
11. English, H. The initial phase of the process of abstraction, *Amer. J. Psychol.*, 33, 1922, 305–350.
12. English, H. A student's dictionary of psychological terms. Yellow Springs, Ohio, Antioch College Press, 1928.
13. Fisher, R. A. Statistical methods for research workers. London, Oliver and Boyd, 1928.
14. Fisher, S. C. The process of generalizing abstraction, *Psychol. Monog.*, 21, 1916, No. 2, 1–213.
15. Gengerelli, J. A. Mutual interference in the evolution of concepts, *Amer. J. Psychol.*, 38, 1927, 639–646.
16. Gray, J. S. A behavioristic interpretation of concept formation, *Psychol. Bull.*, 38, 1931, 65–73.
17. Hilliard, G. H. Probable types of difficulties underlying low scores in comprehension tests, Thesis (Univ. of Iowa), 1922.
18. Hull, C. L. Quantitative aspects of the evolution of concepts, *Psychol. Monog.*, 28, 1920, No. 1.
19. Jacobson, E. On meaning and understanding, *Amer. J. Psychol.*, 22, 1911, 553–577.
20. Jorgensen, A. N. Iowa silent reading examination, Thesis (Univ. of Iowa), 1927.
21. Jung, C. G. Studies in word association (Tr. by Eder). London, Heinman, 1918.
22. Kakise, H. Conscious concomitants of understanding, *Amer. J. Psychol.*, 22, 1911, 14–64.
23. Means, M. H. A tentative standardization of a hard opposites test, *Psychol. Monog.*, 30, 1921, No. 1.
24. Moore, T. V. The process of abstraction: an experimental study, *Univ. of Calif. Publ. in Psychol.*, 1, 1910, No. 2, 1–124.
25. Ogden, R. M. Some experiments on the consciousness of meaning, Studies in Psychology, Titchener Commemorative Volume, Worcester, Louis N. Wilson, 1917. Pp. 79–121.
26. Piaget, J. Judgment and reasoning in the child. New York, Harcourt Brace, 1928.
27. Pressey, L. C. Specific elements making for proficiency in silent reading when general intelligence is held constant, *School and Soc.*, 24, 1926, 589–592.
28. Ribot, T. The evolution of general ideas (Tr. by Welby). Chicago, Open Court Publ., 1899.
29. Rignano, E. The psychology of reasoning. New York, Harcourt Brace, 1923.
30. Rosanoff, A. J. Manual of psychiatry. New York, Wiley, 1920.
31. Schwiete, F. Ueber die psychischen Repräsentation der Begriffe, *Arch. f. d. ges. Psychol.*, 19, 1910, 475–544.

32. SMOKE, K. L. Objective study of concept formation, *Psychol. Monog.*, 42, 1932, No. 4.
33. STODDARD, G. D. Iowa Placement Examinations, *Univ. Iowa Stud.*, 3, 1925, No. 2.
34. THORNDIKE, E. L. Reading as reasoning: a study of mistakes in paragraph reading, *J. Educ. Psychol.*, 8, 1917, 323–332.
35. WEISS, A. P. Meaning as response, Ninth International Congress of Psychology, Proc. and Papers. Psychol. Rev. Co., Princeton, N. J., 1930. Pp. 479–480.
36. WELLS, F. L. The question of association types, *Psychol. Rev.*, 19, 1912, 253–270.
37. WHEELER, R. H. Development of meaning, *Amer. J. Psychol.*, 33, 1922, 223–233.
38. WHIPPLE, G. M. Manual of mental and physical tests. Baltimore, Warwick and York, 1910.
39. WILEY, W. E. Difficult words and the beginner, *J. Educ. Res.*, 17, 1928, 278–89.
40. WOODWORTH, R. S., and WELLS, F. L. Association tests, *Psychol. Monog.*, 13, 1911, No. 5.
41. ROBINSON, F. L. The rôle of eye-movements in reading, Thesis (Univ. of Iowa), 1932.

Appendix

Standardized interview (Series A)

Name...
Date...
Time...
Word...

Definition:

1. When the word was first presented, or while you were formulating your definition, did you see in your mind an outline or image:
   - A. Of the word being defined?
   - B. Of the words you used in your definition?
   - C. Of any other words?
   - D. Of any places, objects, people, or events?
   - E. Were these images:
     - a. Distinct or indistinct?
     - b. Fleeting or long-continued?
     - c. Specific or general?
2. When the word was first presented, or while you were formulating your definition, did you have any auditory images:
   - A. Of the word being defined?
   - B. Of the words you used in your definition?
   - C. Of any other words?
   - D. Of any places, objects, people, or events?
   - E. Were these images:
     - a. Distinct or indistinct?
     - b. Fleeting or long-continued?
     - c. Specific or general?

3. Did you say the word even briefly to yourself as you were defining it?
   Did you say the definition over to yourself as you were formulating it?
   Did you say any other words to yourself?
4. Did you have any sensations or images of movements in any part of the body?
5. Did you have any other ideas that did not involve images of any kind?
6. Did you have any memories:
   A. Of events, places, people, or things?
   B. Of instances when you have seen or heard the word used?
   C. Of instances when you have used the word?
   D. Of any unusual circumstances connected with the word?
   E. Were these memories in the form of images, ideas, or feelings?
7. Did you feel that the word was familiar or unfamiliar when you first saw it?
8. Was there any confusion involved in the process of definition?
9. Did you feel satisfied with the definition that you gave?
10. Did you feel certain that this was a correct definition?
11. Did you feel that you knew a great deal about the word or not very much?
12. Did a large number of ideas come to mind or only a few?
13. Were there any other meanings of the word beside the one given that came to mind as you were thinking of a definition?
14. Was there any feeling of pleasure or displeasure accompanying the formulation of the definition?

Comments:

THE THRESHOLD OF FEELING IN THE EAR IN RELATION TO ARTIFICIAL HEARING AIDS [1]

by

SCOTT N. REGER

In 1922 *Fletcher* and *Wegel* (5) published an unusually complete graph of the auditory sensation area. The essential features of this graph are reproduced below, by permission, as Fig. 1. The ordinates give root mean square pressures in dynes per sq.

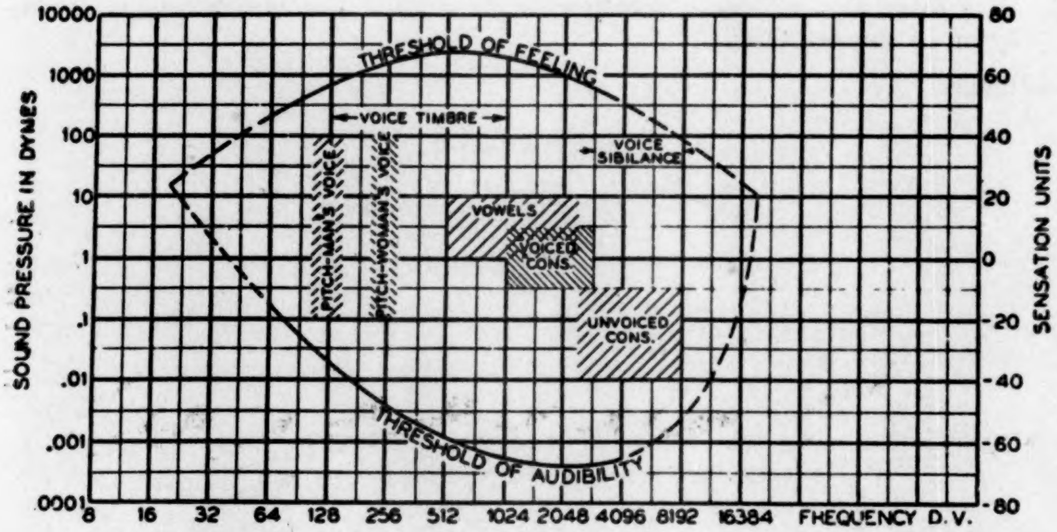

FIG. 1. Any sound that can be heard lies within the field outlined here. Areas covered by the most prominent speech sounds are indicated.

cm. and the abscissæ denote frequency of vibration. Both are plotted on a logarithmic scale. The sensory units on the right hand side of the graph, defined as twenty times the logarithm of

[1] I wish to make acknowledgment to Dean C. E. Seashore for his sympathetic counsel and direction of this study; to D. M. Lierle, M.D., Head of the Department of Oto-Laryngology and Oral Surgery of the University of Iowa Hospitals, in which department the writer holds a Research Assistantship, for his invaluable aid and advice in the diagnosis and selection of the clinical cases involved; and, to Mr. O. L. McIntire, M.A., Superintendent of the Iowa School for the Deaf, Council Bluffs, Iowa, for his coöperation and hospitality while working at that Institution.

the pressure, are quantitative expressions of the intensity of the auditory sensation which the normal ear mediates at any given frequency and sound-pressure within the auditory sensory area. The lower curve of the auditory sensory area, designated as the "threshold of audibility," shows the number of dynes of sound-pressure per sq. cm. necessary to elicit an auditory sensation in 72 normal ears over the indicated frequency range.

The upper curve of the auditory sensory area, designated as the "threshold of feeling," [2] shows the number of dynes of sound pressure per sq. cm. necessary to elicit sensations of tickle, pressure, or pain in 48 normal ears over the indicated frequency range. One of the most significant facts about the auditory sensory area is the finding that the pressures in dynes per sq. cm. necessary to give rise to both the thresholds of audibility and of feeling vary greatly with frequency.

The threshold of feeling of the ear, due to stimulation by excessive sound-pressures, is rapidly assuming more importance in the fields of psychology, otology, acoustics, and the education of the deaf. Owing to its importance, its newness as a concept in the fields of science just enumerated, and the paucity of the literature about it, it seems advisable to quote in chronological order the significant published findings about the threshold of feeling from an historical point of view. Practically all of the work to date has emanated from investigators of the Bell Telephone Laboratories.

[2] The writer is of the opinion that the selection of the phrase, "threshold of feeling," is an unfortunate choice of terminology with which to designate the upper curve of the auditory sensation area. According to *Titchener* (11) 'feeling' is used in psychology technically to denote a simple connection of sensation and affection in which the affection preponderates. Feeling, in the wider meaning of the word, is the general name for the affective side of the mental life. Certain cutaneous experiences, such as hardness, roughness, and solidity, are not feelings at all, but perceptions. Psychologically, therefore, the use of the phrase "threshold of feeling" to denote a threshold of cutaneous sensations may be condemned. However, the originators of the phrase have so defined it that no ambiguity exists as to its meaning in relation to the auditory sensory area. The writer continues the usage of the phrase because it has been employed to date in much of the literature discussing the upper curve of the auditory sensory area, and by redefinition it possesses an exact meaning in this context. Suggested alternatives are: "threshold of extra-auditory sensation," "threshold of cutaneous sensation," or "threshold of maximum audibility."

Fletcher and Wegel (**6**) wrote: "A pressure variation of approximately 1,000 dynes can be felt and produces a sensation of pain. It is practicable to assume that people who experience no auditory sensations at these pressures are totally deaf." *Wegel* (**14**) writes later: "While this point of feeling probably has no relation to the auditory sense it does serve as a practical limit to the range of auditory sensation. A few observations indicate that people with abnormal ears have a point of feeling sound which is not greatly different from that of normal ears, but this, of course, depends on the type of abnormality. The intensity for feeling is about equal to that required to excite the tactile nerves in the finger tips. . . . The feeling sensation in the middle range of frequency is first a tickling sensation and then becomes acutely painful as the loudness is increased. As the frequency is decreased, the sensation of feeling becomes milder until at frequencies around 60 cycles it is sensible as a flutter, but still quite different from the sense of audition. As the frequency is still further decreased to a point where the hearing and feeling lines appear to intersect, it is difficult to distinguish between the sense of hearing and that of feeling. The low point of intersection of the two normal curves of minimum audibility and feeling sense may therefore be taken arbitrarily as the lower tone limit of audibility. For frequencies lower than this it is easier to feel than to hear the air vibration. For any one individual this point can not be determined with accuracy on account of the variation in judgment, but can be determined by extrapolation. A similar intersection of the two curves occurs at some very high frequency."

Fletcher (**2**) says, "When the intensity of a tone is continually increased, a value is reached where the ear experiences a tickling sensation. Experiments show that the intensity for this sensation is approximately the same for various individuals and the results can be duplicated as accurately as those for the minimum intensity value. It was found that if this same intensity of sound is impressed against the finger it excites the tactile nerves. In other words, the sensation of feeling for the ear is practically the same as for other parts of the body. When the intensity goes slightly above this feeling point, pain is experienced. Consequently, this intensity for the threshold of feeling was considered to be the maximum intensity that could be used in any practical way for hearing. The two points where these two curves intersect have interesting interpretations. At these two points, the ear both hears and feels the tone. At frequencies above the upper intersecting point, the ear feels the sound before hearing it, and in general would experience pain before exciting the sensation of hearing. Consequently, the intersection point may be considered as the upper limit in pitch which can be sensed. In a similar way, the lower intersection point represents the lowest pitch that can be sensed."

Fletcher (**3**) writes further: "There is a level of loudness above which sounds begin to affect the sense not of hearing, but of touch, causing a sensation of tickling in the ears which very soon becomes painful. This represents the upper limit of hearing and is called the 'threshold of sensation,' since feeling then begins. It is useless to amplify sounds above this point, for were this done, the sounds would produce pain and not hearing. . . . The threshold of feeling is about the same for all people whether deaf or not. The effect of deafness is to raise or modify in some manner the threshold of audibility."

Jones and Knudsen (**8**) point out: "As the amplitude is increased, the tone becomes louder and louder, until it attains the upper limit of loudness. This upper limit of loudness is reached when a greater increase of amplitude does not increase the loudness sensation. At this loudness level the tone becomes 'painfully loud.' If carried beyond this limit there is not only the sensation of pain

but also the tactile sensation of feeling. For example, even a deaf-mute, though he has no sensation of tone, is able to feel such excessive stimulation."

Banister (1) states: "At a certain intensity, a sensation appears which becomes acutely painful as the loudness is still further increased. It would seem that the latter sensation is extra-auditory. It is observable at approximately the same intensity by abnormal and normal ears. This intensity is about that required to excite the tactile nerves in the finger tips. It is presumed, therefore, that the sensation is tactual, being due to the activation of the nerves of the tympanic membrane, *etc.*"

In 1932 *Wegel* (15), one of the two authors of the original article describing the threshold of feeling, wrote concerning a figure representing a new determination of the threshold of feeling of the ear, "At each frequency, as indicated by the points, the observer was asked to set the intensity first for the auditory threshold and then to increase the sound level until a second threshold was reached. This is usually described as 'feeling,' but its nature varies considerably with frequency and somewhat with observers. At lower frequencies a gentle but definite vibration is experienced which is quite distinct and superimposed on the sound. In some cases, however, 'dizziness' is described, suggesting excitation in the semi-circular canals. At higher frequencies the sensation is one of sharp pain. It is not known which end organ or end organs are responsible for this threshold."

In epitome, the foregoing quotations reveal the following facts bearing significant relationship to the threshold of feeling: (1) when the intensity of tones is gradually increased a level is reached at which the ear responds with a sensation of tickle or pain; (2) the amount of sound-pressure in dynes per sq. cm. necessary to elicit this threshold varies with the frequency; (3) the threshold of feeling is as definite or as critical as the threshold of audibility; (4) the threshold of feeling probably bears no relation to the stimulation of the auditory nerve; (5) approximately the same values of sound-pressure are required to elicit the threshold in normal as in hard-of-hearing ears; (6) the pressure required to elicit the threshold of feeling is about the same as that required to excite the tactile nerves in the finger tips; (7) the quality of the sensation varies with the frequency and the pressure; (8) it is difficult to distinguish between the sense of hearing and that of feeling at the points where the thresholds of feeling and of audibility intersect; (9) these intersecting points determine both the upper and lower limits of audibility in human ears; (10) this threshold serves as a practical upper limit for the intensities of tones which can be perceived by means of the ear on account of the pain which results from increasing the amplification above this value; (11) sound-pressures sufficiently great to elicit the thresh-

old of feeling occasionally stimulate the receptors of the static sense; (12) it is presumed that the sensation is tactual, being due to the activation of the nerves of the tympanic membrane; (13) the exact end-organs responsible for the sensation are unknown.

Pertinent questions may be raised on several issues mentioned in the above quotations. The present paper undertakes to discuss the following: (1) the relationship between the threshold of feeling and the stimulation of the auditory nerve; (2) identity and location of the end-organs responsible for the threshold of feeling; and (3) importance of the threshold of feeling in relation to artificial hearing aids and education of the deaf.

Relationship between the threshold of feeling and stimulation of the auditory nerve. For this aspect of the study the apparatus used consisted principally of the Western Electric No. 2-A audiometer and a high quality two-stage push-pull audio-amplifier. This audiometer generates eight different frequencies: 64, 128, 256, 512, 1024, 2048, 4096, and 8192 ~. The maximum acoustic output at the receiver of any of these eight frequencies is not sufficiently great to elicit the threshold of feeling in normal ears. Therefore the audio-amplifier was utilized further to amplify the electrical output of the audiometer. This amplifier had an overall voltage gain of 48 db and a flat amplification curve (within 2 db) over the range of test-frequencies generated by the audiometer.

As the impedance of the secondary winding of the output transformer of the audiometer is approximately 250 ohms, a 250 ohm impedance line-to-grid input transformer was used to feed the first stage of amplification. The head-phone regularly supplied with the audiometer was connected across a 250 ohm impedance secondary winding on the push-pull output transformer. Attenuation of intensity was controlled by means of the attenuator dial on the audiometer. This dial is calibrated in sensation-units in steps of five units each.

The acoustic output of the receiver, even when loaded with 3.2 watts of power from the amplifier, was insufficient to elicit the threshold of feeling for the three highest frequencies generated by the audiometer with the attenuator dial at maximal setting.

To permit maximum acoustic output of the receiver for the five lowest frequencies, it was necessary to eliminate chattering of the diaphragm against the pole-pieces by increasing the distance between them by means of paper washers. The original damping washers supplied with the receiver were removed. At maximum amplification of the lower frequencies the interaction between the vibrating receiver diaphragm and the receiver case is sufficiently great to elicit a tactile sensation of vibration in the fingers and forearm holding the receiver.

It is assumed that the pressures originating from the spurious frequencies higher than the test-frequency introduced by such wide amplitude of vibration of the receiver diaphragm produced no appreciable error in measurements of the threshold of feeling. Fortunately, the higher the frequency, from 64 to 1024 ~ inclusive, the greater the pressure required to elicit the threshold of feeling.

As the acoustic response of the receiver may not be proportional to the power input at high levels and since equipment for calibrating the acoustic response of it was not available, the results of this study are not based on absolute values of sound-pressures but on relative values and comparisons with the values necessary to reach the threshold of feeling in normal ears. As the writer was not trying to ascertain the absolute values of sound-pressures at the various frequencies necessary to elicit the threshold of feeling, but was interested primarily in the comparative values of the threshold of feeling between normal and abnormal ears, the apparatus served its purpose adequately.

When obtaining readings on the levels necessary to reach the threshold of feeling with the five test-frequencies, the audio-amplifier and modified receiver were utilized to give the required amplification level. The subjects were instructed to inform the investigator when increasing the intensity of each frequency elicited a sensation of tickle, pressure, or pain in the ear. These values were tabulated according to the frequency and the averages secured for each frequency for each group or class of subjects.

Due to the fact that the energy of the stimulus was not increased gradually to the point where the stimulus was just per-

ceptible, but in steps of five sensation-units each, as determined by the calibration of the attenuator dial, the dial readings for any one individual do not present a true picture of the threshold of feeling for this individual. However, the averages of a group of subjects present a reasonably accurate composite picture of the threshold of feeling for the group as a whole.

The attenuator dial settings necessary to elicit the threshold of feeling for the five test-frequencies were secured for 46 normal ears. Ears with a hearing loss not greater than 15 sensation-units for each of the eight frequencies generated by the audiometer alone were considered normal. The phone was held in close contact with the ear for all observations.

Individual variations did not exceed the averages by more than $\pm$ 5 sensation-units in 86 per cent of the subjects. None of the individual variations exceeded $\pm$ 10 sensation-units from the average. The low deviations from the averages are due principally to the fact that sound-pressures slightly above the threshold of feeling produce exceedingly intense sensations.

Comparable data were obtained on 27 subjects with complete bilateral hearing loss. Individuals who experienced no auditory sensations but only pressure, tickle, or pain upon presentation of the auditory stimulus were considered to have complete hearing loss. Sixteen of these subjects lost their hearing because of meningitic infection and 11 were congenital deaf-mutes.[3] These 27 subjects possessed patent external auditory canals and intact tympanic membranes in each ear.

This group of 27 subjects with complete bilateral loss of hearing had normal thresholds of feeling. Their average attenuator dial readings differed less than $\pm$ 5 sensation-units from the parallel averages of the normal ears. Individual differences were no greater than those of the subjects with normal hearing.

The threshold of feeling was next secured for 36 subjects with all degrees of hearing loss and various types of deafness. The 72 ears possessed intact tympanic membranes. Various degrees of hearing loss accompanying different types of deafness gave

[3] Several deaf-mutes with suspected complete hearing loss were found to possess slight amounts of auditory perception.

approximately normal thresholds of feeling. Individual variations were slightly greater ($\pm$ 15 sensation-units below the averages of the group), but the averages differed not more than $\pm$ 5 sensation-units from the parallel averages of the 46 normal ears.

In the light of the above results, the possibility of the sensation of feeling arising from stimulation of the auditory nerve by excessive sound-pressure is exceedingly doubtful. Nevertheless, the possibility of the threshold of feeling being due to stimulation of the auditory nerve by excessively great sound-pressures demands careful consideration.

According to *Herrick* (7): "Some physiologists believe that there are separate nerve-endings for pain; others regard pain as a quality which may be present in any sense, and not as itself a true sensation. . . . The supernormal stimulation of most other sense-organs may result in a very similar type of pain, though in this case the painful quality is accompanied by the normal sensory quality of the organ in question unless the stimulation is excessively strong. From this it would appear that most sensory nerves may upon occasion function as pain nerves. In other cases normal stimulation of a sense-organ may result in a sensation of the quality typical for the organ in question, to which there is added an agreeable or disagreeable quality which may be very pronounced, the disagreeable quality not being painful in the ordinary sense of that term."

Tilney and Riley (10) write: "Each of the several sense-qualities, including touch sense, temperature sense, deep sense, and the distance senses, is capable under certain circumstances of giving rise to sensations quite at variance with that for which it usually serves. Thus, tactile stimuli may, under certain conditions, convey impressions to the brain which, in addition to the fundamental touch perception, provoke a sensation of discomfort, distress, or pain. It would seem that when the stimulus transcends the threshold of mere touch feeling, it tends to reach a level at which it causes actual pain. The significance of this becomes clear when we consider that such stimuli are usually harmful to the tissue. In this manner tactile sense is equipped with an unusual pathway by which impressions unusual in their severity may communicate the nature of their stimulation to the central nervous system which, in turn, responds by an adequate defense reaction. This same accessory protective apparatus is found running parallel with the course of every quality of sensibility. The usual and proper stimuli in each sense quality pass by means of a customary pathway; but when these impressions become excessive and pain results, be it in the sphere of any type of sensation, touch, temperature or muscle sense, then some part of the body is threatened by unfavorable circumstances, and the excessive stimuli follow a special pathway to communicate this information to the central nervous system. This sensory element, forming the basis of a general protective mechanism, is known as pain sensibility (algesthesia)."

These two quotations suggest that extra-auditory sensations may arise from over-stimulation of the auditory nerve. If such be the case, however, it should follow that any condition respon-

sible for raising the threshold of audibility would also simultaneously raise the threshold of feeling. This statement is based on the assumption that (1) a given neurone which is the receptor of a sound stimulus may conduct in addition to its normal sensory quality an extra-auditory sensation when subjected to excessively great stimulation and (2) any condition which would impair the conductivity of a neurone for the transmission of its normal sensory quality would also impair the conductivity of the same neurone for the simultaneous transmission of other sensory qualities.

As previously noted, however, various conditions which raise the threshold of audibility in ears with intact tympanic membranes have no effect upon the threshold of feeling. Furthermore, ears having complete hearing loss due to destruction of the auditory nerve by meningitis and deaf-mutes with complete loss of hearing, still retain normal thresholds of feeling. These findings lead to the conclusion that the threshold of feeling is not causally related in any manner to the stimulation of the auditory nerve.

Identity and location of the end-organs responsible for the threshold of feeling. As previously noted, *Banister* (1) "presumed" that the sensations which define the threshold of feeling are "tactual," being due to the activation of the nerves of the tympanic membrane. The presumption is not supported by experimental or anatomical evidence. The availability to the writer of a wealth of clinical material lacking intact tympanic membranes, both unilaterally and bilaterally, facilitated investigation into this aspect of the study.

The apparatus and technique employed were identical with those described above. The thresholds of feeling for each of the test-frequencies were secured for 33 ears lacking tympanic membranes and for 16 tympanic membranes containing perforations. Eleven tympanic membranes were missing because of middle-ear infection, and 22 had been removed surgically during the radical mastoid operation.

Part of the technique of the radical mastoid operation consists of the removal of the tympanic membrane, malleus, and incus. The malleus is severed from its attachment to the tensor tympani

muscle, which, along with the stapes and stapedius muscle, is left *in situ*. After recovery from the operation, the combined middle ear and external auditory meatus present a lumen lined with cutaneous epithelium. The epithelium lining the former middle-ear cavity or tympanum is somewhat thickened and covers the tensor tympani and stapedius muscles, stapes and round window. The exact degree of damping of the stapes and round window by this epithedial layer is not known but may safely be assumed to be quite high.

The averages of the thresholds of feeling of the 33 ears without tympanic membranes were lowered 20 sensation-units below the parallel averages of the normal ears. Twelve of the ears without membranes did not experience tickle or pain with maximal amplification of the test-frequencies. One subject did not experience the threshold of feeling in an ear possessing an intact tympanic membrane. Teasing this membrane with a blunt metal probe revealed complete absence of the sensation of pressure in this structure. The subjects with unilateral loss of tympanic membranes had normal thresholds of feeling in the ears possessing intact membranes.

The averages of the thresholds of feeling of the 16 tympanic membranes containing perforations were raised approximately 10 sensation-units. The raising of the threshold was roughly proportional to the size of the perforation, although no thresholds were raised more than 10 sensation-units above the normal averages until the size of the perforation exceeded half the area of the tympanic membrane.

The above results strongly suggest that the end-organs responsible for the threshold of feeling are located primarily within the tympanic membrane, at least in ears possessing intact membranes. Any sensations of tickle, pressure or pain in ears without tympanic membranes may be due to the stimulation of contact receptors within the epithelium lining the combined middle-ear cavity and external auditory canal.

Microscopic anatomy and clinical observations of the tympanic membrane support the above conclusion. *Herrick* (7) and *Ranson* (9) believe that free nerve-endings are pain receptors because these endings alone are present in

some parts of the body where susceptibility to pain is the principal sensory quality present, such as the tympanic membrane, central part of the cornea, and dentin and pulp of the teeth.

Wilson (**13**), in addition to the free nerve-endings, has discovered a second type of contact receptor within the tympanic membrane which he describes as follows, " In the fibrous tissue near the periphery there are seen several varieties of endings whose size, shape, and poorly developed capsule enable one to classify as modified vater-pacinian corpuscles. They lie immediately under the epithelium, no papillae being present in this area or in any part of the membrana tympani."

Wilson continues, ". . . as in the cornea, pain and not touch appears to be the sensation evoked, so also in the membrana tympani one might expect that the slighest pressure would evoke unpleasant sensations, passing into pain, a fact well borne out by clinical observations." He then comments relative to the preceding quotation, " Since then I have made many observations to test this hypothesis and find that it is undoubtedly true. By lightly touching the membrana tympani either with a small piece of cotton wool or a fine hair mechanical stimulation from the threshold possesses unpleasantness. . . . if this membrane be touched with a fine hair according as the pressure stroke of the hair is increased, so do we pass from unpleasantness through acute pain in the ear to pain radiating from the n. auriculo-temporalis." *Wilson* ascribed no function other than pain reception to the modified vater-pacinian corpuscles because of the fact that mechanical stimulation of the tympanic membrane resulted in only unpleasant and painful sensations.

It seems probable that if the energy of a mechanical stimulus designed to study the type of cutaneous sensations resulting from stimulation of the tympanic membrane could be controlled from subliminal values to positive sensory values, tickle, pressure, pain, and perhaps other types of cutaneous sensations, could be elicited with pressure values insufficiently great to elicit painful sensations. The energy of the stimuli applied by *Wilson* seems to have been well above the threshold of unpleasantness. It is possible that " tickle " nerve impulses originated from receptors within the tympanic membranes of his subjects but were " masked " by the more intense and unpleasant sensation of pain.

It seems reasonable to assume that the tickle and pressure sensations in the ear due to sound-pressures rise from stimulation of the modified vater-pacinian corpuscles within the tympanic membrane. In the first place, the presence of these corpuscles is not necessary to explain the origin of the sensation of pain resulting from mechanical stimulation of the tympanic membrane. The free nerve-endings of this structure eliminate the necessity of this assumption. Secondly, the stimuli used by *Wilson* in his explora-

tion of the sensitivity of the tympanic membrane to mechanical pressure were not of the type which enabled precise control of their energy output from subliminal to positive sensory values. Third, the presence of a tactile receptor within the tympanic membrane is necessary to explain the origin of the sensations of tickle and pressure when stimulated by sound-pressures. This modified vater-pacinian corpuscle moreover fulfills the requirement.

The unmodified normal pacinian corpuscle is commonly considered to be the contact receptor of deep firm pressure. It is safe to assume that the modification of this corpuscle within the tympanic membrane is the result of evolutionary changes in adapting itself to optimal detection of stimuli most injurious to the particular tissue in which it is located.

Titchener (11) says: "It is significant that the areas which are especially ticklish are highly vulnerable: in nearly all of them some important structure, such as a large artery, is close to the surface; and where this is not the case, as in the sole and the palm, an injury, even if not serious, is seriously disabling to the organism."

The distinction between tickle and touch is quite definite, the former being correlated with a light and unsteady pressure, whereas the latter accompanies firmer contacts. *Troland* (12) writes: ". . . the ticklish quality of touch tends to go with weak stimulation of small areas of the skin. Its evocation is also favored by maximally brief action of the stimulus. . . . The basis for the difference between tickle and touch would appear to lie in the temporal and spatial patterns of the two excitations, tickle being evoked when the process is restricted to a small number of receptors and to a brief interval of time, whereas touch goes with a more extensive and persistent excitation. Tickle would belong in the class of 'phastic' (momentary) and touch in that of 'postural' (relatively persistent) sensory response. It is possible, however, that there is a selection between separate classes of tactile end-organs, such as protopathic *versus* epicritic, on the ground of stimulus intensity difference, the tickle organs lying more superficially in the skin and being excited phastically by pressure changes which distort the tissues only slightly, while the more 'postural' tactual receptors are aroused only by pressures which penetrate more deeply."

The requirements of an adequate stimulus with which to elicit tickle, pressure, and pain sensations in the ear are ideally fulfilled by sound-pressures. The amount of energy exerted upon the receptor can be controlled accurately from subliminal to practically any desired sensory value and the rate of application of the stimulus is varied with the frequency. By means of a General

Radio Company, type 613-A, beat-frequency oscillator and the audio-amplifier and receiver previously described, it is possible to produce sensations of combined pressure and tickle in the ear with a frequency as low as 1 ~. The amplification level at this frequency is not sufficiently high to produce pain. The wave-form of the sound-pressures at this low frequency, of course, can not be represented by a sine wave, and the auditory sensation consists of a discrete series of "puffs" or explosions.

As the frequency is increased the sensation of tickle increases in intensity and the sensation of pressure gradually decreases until at approximately 12 ~ only the sensation of tickle remains. The amplification level can be sufficiently increased at this frequency to produce pain.

As the frequency is gradually increased above 12 ~ and the amplification level held or adjusted to values just sufficiently great to elicit a definite extra-auditory sensation, the tickle sensation is most prominent until a frequency of approximately 1200 ~ is reached. From this point upward, a pressure sensation in the ear, similar to that produced by rapid descent of several floors in an elevator, is most prominent, until at frequencies near 3200 ~ the acoustic output of the receiver falls below the extra-auditory sensory level and produces only an auditory sensation. The acoustic output of the receiver is too low to produce pain with sound pressures having frequencies higher than about 2400 ~.

At levels which produce intense pressure sensations, but insufficiently great to produce pain, several other interesting phenomena are presented. If the stimulus be turned on instantly, a peculiar unlocalized sensation involving a "moving" and a new auditory sensation are elicited simultaneously. The new auditory sensation resembles the sound of a distant waterfall, or the rushing of blood occasionally heard in the ears by some individuals. This sensation does not continue during protracted presentation of the stimulus but is perceived again when the stimulus is turned off instantly.

If the frequency be varied rapidly over the range from approximately 1200–3200 ~ by means of the frequency selector dial on

the beat-frequency oscillator at levels which produce this new auditory sensation, a peculiar " whirring " noise is perceived. The investigator at present holds the view that these phenomena are due principally to the sensations produced by contraction of the tensor tympani muscle. Further discussion of these phenomena is reserved for a future paper. At the frequencies and pressures necessary to elicit the sensation of pressure, stimulation of the semi-circular canals also usually occurs.

Analysis of the acoustic principles involved in the production of the sensations of tickle, pressure, and pain proved significant. The tickle sensation apparently is due to the recurrent displacements of the tympanic membrane from its normal position at a rate coinciding with the frequency of the sound-pressure stimulus. The force exerted by a positive sound-pressure displaces the membrane medially and the negative pressure displaces it laterally. The forces tending to restore it to its resting position consist of the atmospheric pressures on each side of the tympanic membrane plus the elastic forces in the membrane and its associated structures.

The pressure sensations, resulting from stimulation with the higher frequencies, are due to practically permanent displacements of the tympanic membrane medially because of its reflection of the variations in sound-pressure. The inertia of the tympanic membrane and associated structures is such that at these frequencies the slight vibratory movements of the membrane, due to the individual pressure-waves, produce only subliminal tickle stimuli.

The sensation of pain is due to the tension exerted on the tympanic membrane by excessively great sound-pressures. The pain is very acute in an ear with a bulging tympanic membrane due to the pressure exerted by the accumulation of liquid in the middle ear because of infection. Spontaneous rupture or myrngotomy of the membrane instantly relieves the pain. It is well known that pain in the intestine is elicited by tensions—stretching and pulling—rather than by cutting or pinching. The raising of the threshold of feeling in ears with perforated tym-

Radio Company, type 613-A, beat-frequency oscillator and the audio-amplifier and receiver previously described, it is possible to produce sensations of combined pressure and tickle in the ear with a frequency as low as 1 ~. The amplification level at this frequency is not sufficiently high to produce pain. The wave-form of the sound-pressures at this low frequency, of course, can not be represented by a sine wave, and the auditory sensation consists of a discrete series of " puffs " or explosions.

As the frequency is increased the sensation of tickle increases in intensity and the sensation of pressure gradually decreases until at approximately 12 ~ only the sensation of tickle remains. The amplification level can be sufficiently increased at this frequency to produce pain.

As the frequency is gradually increased above 12 ~ and the amplification level held or adjusted to values just sufficiently great to elicit a definite extra-auditory sensation, the tickle sensation is most prominent until a frequency of approximately 1200 ~ is reached. From this point upward, a pressure sensation in the ear, similar to that produced by rapid descent of several floors in an elevator, is most prominent, until at frequencies near 3200 ~ the acoustic output of the receiver falls below the extra-auditory sensory level and produces only an auditory sensation. The acoustic output of the receiver is too low to produce pain with sound pressures having frequencies higher than about 2400 ~.

At levels which produce intense pressure sensations, but insufficiently great to produce pain, several other interesting phenomena are presented. If the stimulus be turned on instantly, a peculiar unlocalized sensation involving a " moving " and a new auditory sensation are elicited simultaneously. The new auditory sensation resembles the sound of a distant waterfall, or the rushing of blood occasionally heard in the ears by some individuals. This sensation does not continue during protracted presentation of the stimulus but is perceived again when the stimulus is turned off instantly.

If the frequency be varied rapidly over the range from approximately 1200–3200 ~ by means of the frequency selector dial on

the beat-frequency oscillator at levels which produce this new auditory sensation, a peculiar "whirring" noise is perceived. The investigator at present holds the view that these phenomena are due principally to the sensations produced by contraction of the tensor tympani muscle. Further discussion of these phenomena is reserved for a future paper. At the frequencies and pressures necessary to elicit the sensation of pressure, stimulation of the semi-circular canals also usually occurs.

Analysis of the acoustic principles involved in the production of the sensations of tickle, pressure, and pain proved significant. The tickle sensation apparently is due to the recurrent displacements of the tympanic membrane from its normal position at a rate coinciding with the frequency of the sound-pressure stimulus. The force exerted by a positive sound-pressure displaces the membrane medially and the negative pressure displaces it laterally. The forces tending to restore it to its resting position consist of the atmospheric pressures on each side of the tympanic membrane plus the elastic forces in the membrane and its associated structures.

The pressure sensations, resulting from stimulation with the higher frequencies, are due to practically permanent displacements of the tympanic membrane medially because of its reflection of the variations in sound-pressure. The inertia of the tympanic membrane and associated structures is such that at these frequencies the slight vibratory movements of the membrane, due to the individual pressure-waves, produce only subliminal tickle stimuli.

The sensation of pain is due to the tension exerted on the tympanic membrane by excessively great sound-pressures. The pain is very acute in an ear with a bulging tympanic membrane due to the pressure exerted by the accumulation of liquid in the middle ear because of infection. Spontaneous rupture or myrngotomy of the membrane instantly relieves the pain. It is well known that pain in the intestine is elicited by tensions—stretching and pulling—rather than by cutting or pinching. The raising of the threshold of feeling in ears with perforated tym-

panic membranes is undoubtedly due to the rapid equalization of sound-pressures on each side of the membrane that is facilitated by the perforation.

It is interesting to note that although the sound-pressures required to elicit the threshold of feeling in the ear are approximately equal to those required to produce a tactual sensation in the finger tips, the qualities of the sensations at these two diverse locations are not similar. A pressure which just produces tickle in the ear is perceived at the finger-tips as a very light " touch " or contact sensation. In some instances, depending upon the mode of stimulation, the sound-pressure is interpreted as a current or jet of air blowing against the finger. Pressures which cause intense pain in the ear do not elicit even faintly unpleasant sensations at the finger-tips.

Although the experimental work described above on ears with and without tympanic membranes seems to point conclusively to the location of the receptors responsible for the normal threshold of feeling within the tympanic membrane, other possible locations must not be overlooked. As shown above, the threshold of feeling is not due to excessive stimulation of the auditory nerve. Other than the auditory nerve and tympanic membrance, the possibility of the threshold of feeling arising from stimulation of receptors within the saccule and utricle, round window, tensor tympani and stapedius muscles, and articulations of the ossicles seems most plausible.

Seven of the cases with complete hearing loss but with normal thresholds of feeling also had lost all vestibular reaction due to destruction by meningitis of the vestibular portion of the VIIIth nerve, as determined by whirling chair and caloric tests. This complete destruction of the VIIIth nerve eliminates the possibility of any sensation arising from receptors of the saccule, utricle, or semicircular canals.

As described above, perforations of the tympanic membrane raise the threshold of feeling about 10 sensation-units. These perforations permit the exertion of greater sound-pressures directly against the round window, articulations of the ossicles,

tensor tympani and stapedius muscles. This fact is interpreted to indicate that the threshold of feeling is not due to air-pressure stimulation of these structures transmitted through the tympanic membrane to the air within the middle ear.

The threshold of feeling in relation to artificial hearing aids and education of the deaf.

Fletcher and *Wegel* (5) write, ". . . It is sufficient here to give from our general experience the amount of sound volume in the speech range that is required to make people of various degrees of deafness hear. Persons who have normal hearing require approximately 1/1000 dynes per sq. cm. in order to hear sounds in this range. Persons who require a pressure variation of 1/10 dynes per sq. cm. are called slightly deaf. Those who require one dyne are partially deaf, but can usually follow ordinary conversation. Those who require 10 dynes belong to that class who use ear trumpets or deaf sets to amplify the speech waves. A pressure variation of approximately 1,000 dynes can be felt and produces a sensation of pain. For practical purposes, it may be assumed that people who experience no auditory sensation at these pressures are totally deaf. This shows that among people who can follow ordinary conversation there is a range in ear sensitivity of more than 1,000 and among people who are noticeably deaf there is another range of 1,000, making a total range of more than a million for people who can hear or be made to hear by means of amplifying devices."

Fletcher (4) states that for individuals with normal hearing the variations of the intensity of speech, with the ear 1 m. from the mouth of the speaker, include a range from about 40 to 70 db.[4] The average noise level in which conversation usually occurs is approximately 30 db, the faintest sounds heard being about 10 db above this noise level. An individual with a hearing loss of 30 db would not hear the noise, but would experience no difficulty in understanding speech. An individual with a hearing loss of 70 db or more would not hear the speech at all. "Bringing the mouth of the speaker from one meter to close to the ear of the listener will increase the speech intensity about 40 db. This indicates that a person with as much as an 80 db loss of hearing can hear and understand speech when it is spoken directly into the ear."

Fletcher (4) writes that persons with hearing losses between 80 to 110 db require powerful hearing aids involving vacuum tubes and can not understand speech at all without such equipment. Individuals with hearing losses greater than 110 db can not use such equipment. "The vowel sounds hurt them before the consonant sounds are above their threshold."

A very large percentage of students in educational institutions for the deaf have 110 db hearing losses. Individuals with congenital or early acquired hearing losses of this extent have not learned to speak because of their inability to hear the speech sounds of themselves or others. Their attempts to use vacuum

[4] The term 'decibel,' here abbreviated db, is synonymous with 'sensation-unit.'

tube equipment results in pain and disappointment. The threshold of feeling is elicited by the energy in the vowel sounds before the energy of the consonant sounds stimulates the threshold of audibility. As previously pointed out, the effect of deafness is to raise the threshold of audibility, not affecting the threshold of feeling. Also, levels slightly above those necessary to elicit the threshold of feeling result in intense pain. If, in some manner, the threshold of feeling of an ear with a 110 db hearing loss could be raised, the range of amplification which this ear could utilize for sound perception would be correspondingly increased.

The experimental work on the threshold of feeling of ears with and without tympanic membranes shows that the threshold of feeling of ears lacking membranes is raised approximately 20 db above the threshold of ears possessing intact membranes. Furthermore, increasing the energy of the stimulus several db above the values necessary to elicit the threshold of feeling in ears without membranes and 110 db hearing loss does not result in the intense pain characteristic of an equal increase in the stimulation of ears possessing intact membranes. It was impossible to elicit a sensation of unpleasantness in some of the subjects lacking membranes by greatly overloading the capacity of the acoustic output of the receiver.

The results of the work on the threshold of feeling on ears with and without tympanic membranes suggest that removal of the membranes of individuals with 110 db hearing loss would sufficiently raise the threshold of feeling to permit stimulation by the sound-energy levels necessary to produce perception of speech sounds. Of course, auditory stimuli can not be produced in an ear in which the hearing loss is complete, due to congenital defect or complete degeneration of the receptors within the cochlea, the auditory nerve, its tracts, or centers.

Surgical removal of tympanic membranes of individuals with 110 db hearing loss would not raise their threshold of audibility. This statement is based on unpublished results of hearing tests secured on patients both before and after the radical mastoid

operation. Audiometric tests were continued for a period of two years at six month intervals after epithelialization of the middle ear cavity.

Intensive study of one subject with a 110 db hearing loss bilaterally, and bilateral loss of tympanic membranes lends further support to the present issue. This subject, 48 yrs. old, lost all audibility of the spoken voice of others within a time interval of two years following severe middle ear infection coincident with scarlet fever at 14 years of age. He has not heard his own voice, except when forced, for about 30 yrs. The intensity of his speaking voice is controlled principally by means of kinaesthetic sensations arising from his vocal mechanism. Bone conduction tests elicit only sensations of vibration.

This individual has high intelligibility for speech by means of equipment consisting of a condenser type transmitter, audio-amplifier and receiver. He hears his own voice distinctly by means of this equipment even when he is from 10–12 ft. away from the transmitter. He was much amused at hearing his voice in this manner and reported that the situation seemed to make two individuals of him, the one talking to the other. Acoustic feedback between the receiver and microphone of this particular set-up limits the nearness with which these input and output devices can be placed in relation to each other.

A receiver was attached, with due regard to the matching of output and input impedances and current carrying capacity of the receiver, to a high quality radio. The level which suited this individual for highest intelligibility of speech was sufficiently high to cause intense pain in the ears of the investigator. The only extra-auditory sensation this individual perceived was an occasional air-current, the pressure waves of the sound, in the external auditory canal. Occasionally tactual sensations of vibration were elicited from the side of the head and the fingers holding the receiver, due to the stimulus produced by the interaction between the receiver-diaphragm and its case.

The writer is building a hearing aid specially designed to supplement most effectively the hearing requirements of this

individual's threshold of hearing over the frequency-range most essential for high intelligibility of speech sounds. This individual's speech, although intelligible, is sadly lacking in inflection and "clean cut" diction. The equipment will be used by him also as a means of improving and "monitoring" his own speech.

Raising the threshold of feeling to the extent indicated by the absence of tympanic membranes of individuals with bilateral 110 db hearing loss (but not complete loss), plus intelligent use of properly designed amplification equipment, will benefit the following classes of such individuals in the manner to be designated:

1. Individuals who had learned speech before occurrence of the hearing loss. These persons will be able to hear and understand speech more or less adequately, depending upon the degree of hearing loss and suitability of the equipment to meet their hearing requirements. They will be able to regain and maintain more normal inflection and diction of their own speech by utilizing the equipment to hear their own voices.

2. Individuals who never learned speech because of their hearing defect. By means of proper equipment these persons will be able to hear speech sounds of others and of themselves and by a carefully planned regime of training will learn speech more efficiently from the points of view of time and naturalness, than by any other method. Both of the above groups of hard-of-hearing individuals will benefit greatly by correlating sight and sound in lip reading study.

The writer wishes to point out emphatically that the present paper does not advise surgical removal of the tympanic membrane to permit use of higher amplification of sound by exceedingly hard-of-hearing individuals. All such implications are purely theoretical, based on the experimental results described. The only conservative practical application apparent at the present time is the enlarging of perforations already present. The clinical and experimental results of enlarging perforations of long duration and removal of tympanic membranes only when warranted by the extent of middle ear infection, to enable exceedingly hard-of-hearing individuals to use high levels of sound amplification, will

be described in a forthcoming paper by D. M. Lierle, M.D., Head of the Department of Oto-Laryngology and Oral Surgery at the University of Iowa Hospitals, and the writer.

Conclusions:

1. Individuals with complete bilateral hearing loss and individuals with all types and degrees of deafness, but with intact tympanic membranes, possess normal thresholds of feeling.
2. The threshold of feeling is not due to over-stimulation of the auditory nerve.
3. The threshold of feeling of ears lacking tympanic membranes is greatly raised and the threshold of feeling of tympanic membranes containing perforations is appreciably raised.
4. The threshold of feeling of normal ears is due to the stimulation of receptors within the tympanic membrane.
5. The threshold of feeling of ears lacking tympanic membranes is due to the stimulation of receptors within the epithelial lining of the middle ear cavity and external auditory meatus.
6. Individuals with 110 db hearing loss, but not with complete loss, are able to utilize sufficient amplification to permit perception of speech sounds, provided their tympanic membranes are missing.

Bibliography

1. Banister, H. Hearing I. The foundations of experimental psychology. Ed. by C. Murchison, Worcester, Clark Univ. Press, 1929, 907.
2. Fletcher, H. Physical measurements on audition and their bearing on the theory of hearing, *Bell Syst. Tech. J.*, 2, 1923, 145–180.
3. Fletcher, H. Hearing aids and deafness, *Bell Lab. Record*, 5, 1927, 33–37.
4. Fletcher, H. Can we scientifically advise patients as to the effectiveness of hearing aids? *Ann. Otol., Rhinol. & Laryngol.*, 41, 1932, 727–739.
5. Fletcher, H., and Wegel, R. L. The frequency-sensitivity of normal ears, *Phys. Rev.*, June, 1922, 553–565.
6. Fletcher, H., and Wegel, R. L. The frequency-sensitivity of normal ears, *Proc. Natl. Acad. Sc.*, 8, 1922, 5–7.
7. Herrick, C. J. An introduction to neurology. 4th Ed. Philadelphia, Saunders, 1928, 406.
8. Jones, I. H., and Knudsen, V. O. Facts of audition, *Ann. of Otol., Rhinol. & Laryngol.*, 34, 1925, 1013–1027.
9. Ranson, S. W. The anatomy of the nervous system (Md. Ed.). Philadelphia, Saunders, 1925, 421.

10. Tilney, F., and Riley, H. A. The form and functions of the central nervous system (Md. Ed.). N. Y., Paul B. Hoeber, 1923, 1019.
11. Titchener, E. B. A text-book of psychology. N. Y., Macmillan, 1919, 565.
12. Troland, L. T. The principles of psychophysiology. Vol. II, Sensation. N. Y., Van Nostrand, Inc., 1930, 397.
13. Wilson, J. G. The nerves and nerve-endings of the membrana tympani of man, *Amer. J. Anat.*, 11, 1911, 101–112.
14. Wegel, R. L. The physical examination of hearing and binaural aids for the deaf, *Proc. Natl. Acad. Sc.*, 8, 1922, 155–160.
15. Wegel, R. L. Physical data and physiology of excitation of the auditory nerve, *Ann. Otol., Rhinol. & Laryngol.*, 41, 1932, 740–779.

THE EYE-MOVEMENTS OF GOOD READERS

by

ROBERT Y. WALKER

Introduction. This study, confined entirely to the eye-movement phenomena of good readers, forms another link in the general program of research being conducted by the University of Iowa laboratories.[1] Previous studies undertaken in this laboratory have dealt with the eye-movements of stutterers and poor readers.

Erdmann and Dodge (4), *Huey* (7), and *Dearborn* (2) have studied some of the variations of eye-movements for objective factors of reading by varying the length of line of the reading material. *Gilliland* (5) and *Tinker and Paterson* (22, 23) have noted the variations resulting from changes of type-size and type-face. The variation of eye-movement with age and maturity of reader has been examined by *Schmidt* (17), *Gray* (6), and *Buswell* (1). These experimenters have used subjects ranging in age from six to eighteen years and over. None of these workers has had groups of any particular age large enough for statistical treatment of results. As has been shown by *Gray* (6) and *Judd and Buswell* (12), eye-movements vary with the purpose of the reading. No actual study has been undertaken of the movements of the eyes in scanning or skimming. *Judd and Buswell* (12) also studied the variations of the eye-movements with the nature of the reading material. They included different subject matter in the native language and some selections in foreign languages. *Tinker* (20), working with prose, and mathematical and chemical formulæ, has pointed out the differences that exist in the eye-movements for these two types of reading materials. The motor habits of eye-movements have been studied by *Dearborn* (2), *Huey* (7), *Judd and Buswell* (12), and *Tinker* (20). They found evidence indicating the existence of short-lived motor habits that are dependent upon the nature of the material being read. The eye-movements of poor readers have been examined by *Robinson* (15). The subjects were trained in a reading clinic for the purpose of determining the eye-movement factors and their relation to poor reading and to study the effect of improvement of habits upon the eye-movements. *Jasper and Murray* (9) have conducted investigations into the eye-movements of stutterers, finding evidence of "blocks" similar in nature to those that exist in the speech-muscles. The works of *Robinson* (15) and *Jasper and Murray* (9) pointed to the need of a thorough investigation of

[1] The writer wishes to acknowledge his indebtedness to Dean Carl E. Seashore for his suggestion of the problem and for his critical guidance in its execution; and to Dr. Lonzo Jones, Assistant Dean of Men, for his aid in securing students through the authority of his office.

the eye-movements of good readers. Such a study would fill out the field for clinical purposes.

Preliminary examinations were made of the records of ten subjects in order to discover related problems that might be incorporated in this study. From the results of this preliminary examination, the following program was set up: (1) to analyze the eye-movements and related phenomena of good readers in such a manner as to be statistically reliable, (2) to study the effect of change of difficulty of reading material upon eye-movement phenomena, (3) to observe the effect on eye-movements of alteration of comprehension, (4) to establish norms for the eye-movements of good readers, (5) to compare the relationship between span of fixation and span of perception, (6) to establish the reliability and validity of the objective methods used in this study.

Experimental procedure

Subjects. The subjects for this study were university freshmen from the academic years 1930–31 and 1931–32, selected on the basis of their scores obtained on the *University of Iowa Qualifying Examination* and *Iowa Silent Reading Test.* All subjects ranked in the highest decile for both tests for all freshmen students. Through the facilities of the Dean of Men's office, the subjects were called in for a preliminary examination. Inasmuch as the corneal reflection method of eye-photography does not permit the reader to wear glasses during the recording, readers with emmetropic, or near emmetropic vision, were chosen.

Reading material. Six paragraphs of reading material were selected for the purpose of this study. Selection II was used for an analysis of the eye-movements and for setting up norms of the eye-movements; I, II, and III were used to study the effect of change of difficulty of material; IV, II, and V introduced change of comprehension requirements; and VI was designed for the study of skimming. A description of each paragraph appears under the selection in which it was used. All paragraphs were set up in 10 point DeVinne type 25 ems (106 mm.) to the line.[2]

[2] Samples of all reading paragraphs are contained in *Appendix A,* which is included in the Thesis on file in the Library of the University of Iowa.

For I, II, III, and VI a comparable paragraph was chosen for use as an introductory exercise in order to familiarize the subject with the requirements of this part of the study in the experimental situation. Comprehension was checked after the reading of each paragraph.

Span of perception test. Four groups, each of four, five, six, and seven words, were cut from each of I, II, and III. These word groups were mounted on cards and presented to the subject in a *Whipple* tachistoscope. This test preceded the recording of the eye-movements.

Eye-movement records. Records of eye-movements were obtained by means of the *Iowa Eye-Movement Camera* (**10**). This camera, which utilizes the corneal reflection technique of photography, represents a marked advance in that it permits simultaneous binocular recording of both the vertical and horizontal eye-movements. A strong beam of light from an automatic carbon arc lamp was passed through a system of condensing lenses and a cobalt blue filter to a pair of first surface mirrors. The light was reflected from the mirrors to the cornea of each eye.

In order to obtain simultaneous binocular records for both the vertical and horizontal movements, the beam of light reflected from the cornea of each eye was split into two beams by prisms mounted on the focussing lenses of the camera. One beam was focussed upon a sensitized paper moving in a vertical direction, the other beam upon a sensitized paper moving in the horizontal direction. Both the right and left lenses of the camera could be adjusted independently for differences in width, elevation, or depth between the two eyes, so that the records from the right and the left eye would always be parallel. Time in units of .02 sec. was recorded by interrupting the beam of light with a segmented revolving disc operated by a synchronous motor. Eastman No. 1 recording paper 1–7/8" in width was used in place of the ordinary moving-picture film. This paper costs about one-fourth as much as the film, permits a wider record, and is as sensitive as panchromatic film. The sensitized paper was fed through the recording box by toothed rollers and travelled synchronously, so that they could be matched at any time. For the localization of syn-

chronous points, an auxiliary time-marker was mounted in the recording box so that a beam of light could be directed upon both sensitized papers at their point of crossing.

The sample record in Fig. 1 shows the horizontal and the vertical movements of the eyes in reading a line of printed material. The dots represent .02 sec., the bars appear every .5 sec., and .78 mm. movement on the record represents 1 degree of angular eye-movement. The first horizontal movement in Record A is that of the eyes moving from the end of one line to the first of the succeeding line. This is represented in B by a slight overlap of the dots. The second and third horizontal movements in A are forward shifts. The fourth movement is a regression. The spreading of the dots in B represents forward shifts; the short overlap, the regression. The overlaps are caused by a movement of the eye in the same direction that the sensitized paper is traveling, but since the eye-movement is faster than the movement of the paper, the records are superimposed.

The reading material was mounted on an easel 15 inches in front of and slightly below the eyes of the subject. A screen covered the paragraph until all focussing had been completed. Immediately above the first and last letter of the first line of the printed material were two fixation points so placed that the screen did not cover them. At the start of each recording, the subject was asked to fixate alternately upon these points until the screen was removed from the paragraph, when he should immediately begin reading. The records of these fixations were used as marginal limits for determining the width of the line on the record and as a basis from which the size of the shifts in the line was determined.

The subject was instructed to read through the paragraph once, to make necessary regressions within any one line, but not to return to a previous line, and upon completion of the paragraph to close his eyes. Every possible attempt was made to put the subject at ease and to overcome the strangeness of the recording situation. Subjects were requested to read in a normal manner. During the focussing of the camera and the presentation of the preliminary paragraph, most of the apparatus was in operation, so that the subject could become accustomed to the noise. Between each recording the subject was permitted to remove his head from the head-rest and to relax. The only objection made to the experimental situation by the subjects was the intensity of the beam of light directed on the eyes. However, if they looked

only at the reading material, the beam of light fell upon the periphery of the retina and, since the beam was passed through a cobalt filter, it was not sufficiently strong to cause much annoyance. The filter reduced the white glare that might annoy the subject but permitted the actinic rays to pass through and activate the sensitized paper.

Unit of measurement: the em. The em was used as a unit of reading material constituting a linear form of measurement. All spaces between words and sentences had to be included in any measure and therefore the em becomes a fixed fraction of a line. This tends to reduce the number of ems in any word, because

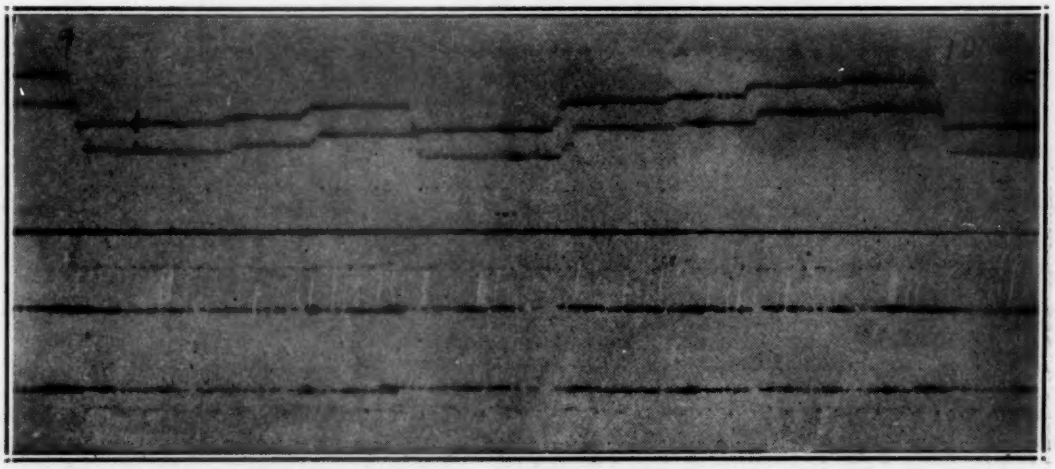

FIG. 1. Record A is the vertically moving sensitized paper which records the horizontal movements of the eyes. Record B is the horizontally moving paper which records the vertical movements of the eyes.

the number of ems per word was calculated by dividing the total em length of a paragraph by the total number of words. This method of measurement resulted in the number of ems per word, varying from 1.97 to 2.28 for the different paragraphs with an average of 2.16. The average given by linotypers is 2.17 ems per word.

Measurement of records. In measuring the shift in the line, a pantograph was adjusted for the marginal limits of the record as determined by the fixations on the dots above the first and last letter of the first line. The size of each shift was measured by reading the location in the line of each fixation. Individual differences in the distances between the marginal limits were cor-

rected by the use of a sliding scale under the pointer of the pantograph.

Tables of Group Data[3]

Table I. *Seven different measures on the six types of reading material*

Selection	I	II	III	IV	V	VI
A	.246	.255	.277	.240	.285	.257
	.035	.033	.035	.026	.040	.030
B	.091	.101	.108	.086	.119	.095
	.035	.032	.025	.020	.037	.027
C	.220	.213	.233	.216	.233	.223
	.035	.033	.041	.046	.035	.038
D	3.40	2.92	2.66	3.75	2.52	—
	.54	.60	.43	.67	.48	—
E	4.31	4.31	4.06	4.55	3.91	4.35
	.75	.64	.56	.80	.52	.80
F	.44	.70	.92	.33	1.02	.60
	.06	.06	.12	.29	.66	.51
G	758.2	690.9	558.1	882.4	513.3	769.2
	130.0	110.5	101.8	169.2	127.9	152.4

Legend:

A. Mean duration per fixation.
B. Mean of the individual S.D.'s of the duration per fixation.
C. Mean mode of duration per fixation.
D. Mean size of fixation in ems.
E. Mean extent of forward shift of eye in ems.
F. Mean regressions per line.
G. Mean rate of reading in ems per min.
In the tables, the S.D. for each measure is placed immediately under the mean.

Table II. *The critical ratio (C.R.) for the differences existing between any two means of the six selections for seven measures of eye-movements*

Measure	A	B	C	D	E	F	G
I & II	1.8	1.5	1.9	3.6	.03	2.9	6.6
II & III	4.0	1.5	3.3	3.2	3.1	2.5	6.8
I & III	3.2	3.6	2.4	7.0	3.1	3.8	11.9
IV & II	3.6	1.4	.7	8.7	2.1	3.7	6.8
II & V	4.1	3.5	2.5	5.3	3.1	5.6	5.4
IV & V	7.4	6.1	1.2	12.3	4.7	5.0	10.7

Legend same as in Table I.

$$\text{C.R.} = \frac{M_1 - M_2}{\text{S.D. difference}}$$ between any two measures.

Analysis of Eye-Movements for Moderate Difficulty and Comprehension

Selection II is of moderate difficulty, comparable to the average reading material that is encountered in college reading. The requirement of comprehension, which was a moderate knowledge

[3] Tables for the individual data on this study are contained in *Appendix B* which is included in the Thesis on file in the Library of the University of Iowa.

of the contents, was checked at the completion of reading by demanding three correct answers to four questions. Answers to two questions were contained in the paragraph; answers to the other two were to be inferred from the paragraph.

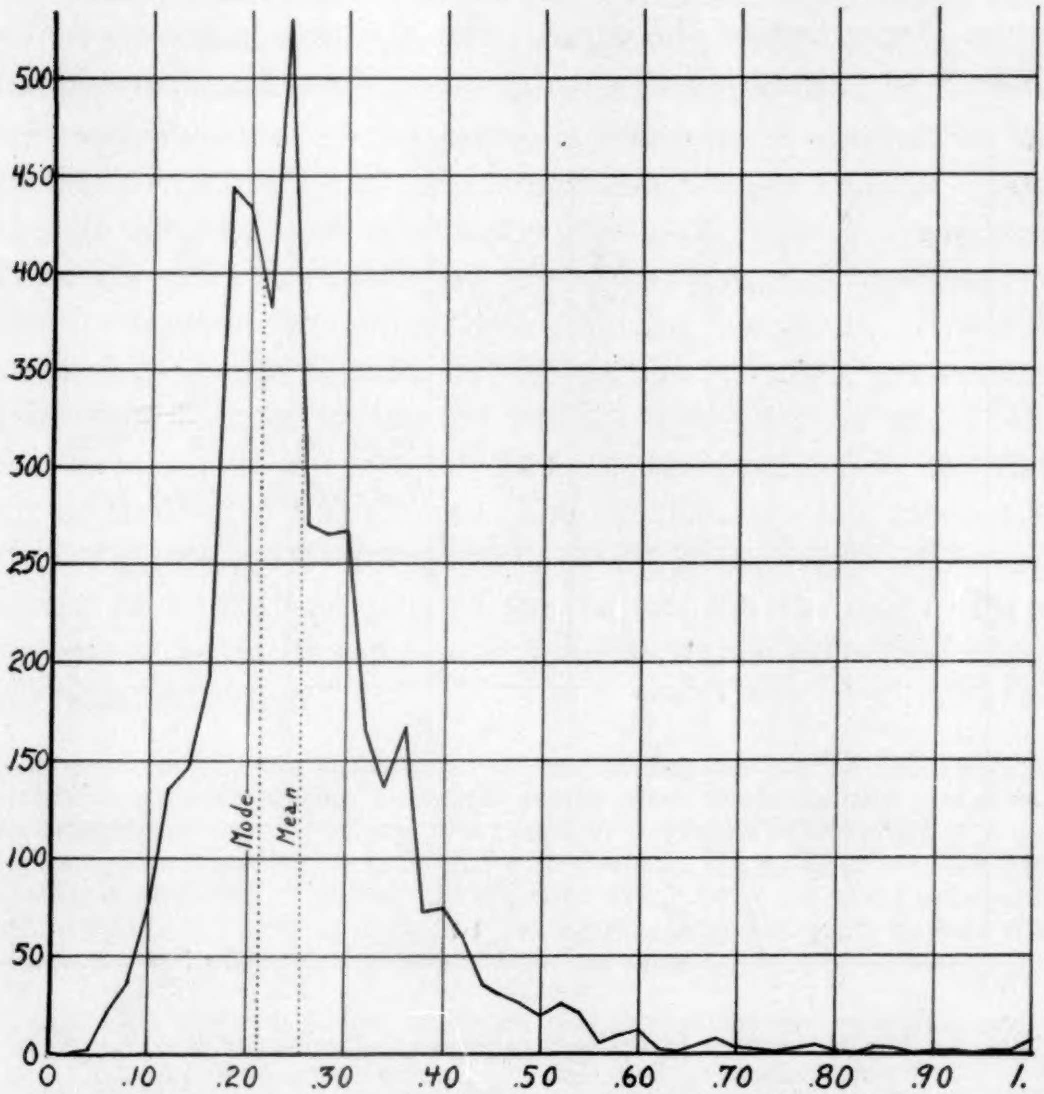

FIG. 2. The distribution of the individual duration per fixation of 50 good readers for Selection II. Time in sec. is plotted on the abscissae; frequency on the ordinates. To determine the mode the curve was smoothed by the method of rolling averages.

Fixations. The mean duration for all subjects in this study was found to be .255 sec.; S.D., .033 sec. These figures indicate that for good readers the individual mean for any reader should fall within .16 to .34 sec. However, such figures do not show the variability of the individual. The mean of the individual

S.D.'s for the group is .10 sec., S.D., .031 sec. Thus an individual might have a much greater variability than the group.

As shown in Fig. 2, the curve of distribution for the entire group as regards duration of fixation is highly skewed, with a large grouping at the shorter end and a rather pronounced spread at the longer end of the curve. For this reason the modes of duration of fixation were plotted, because it was felt that although

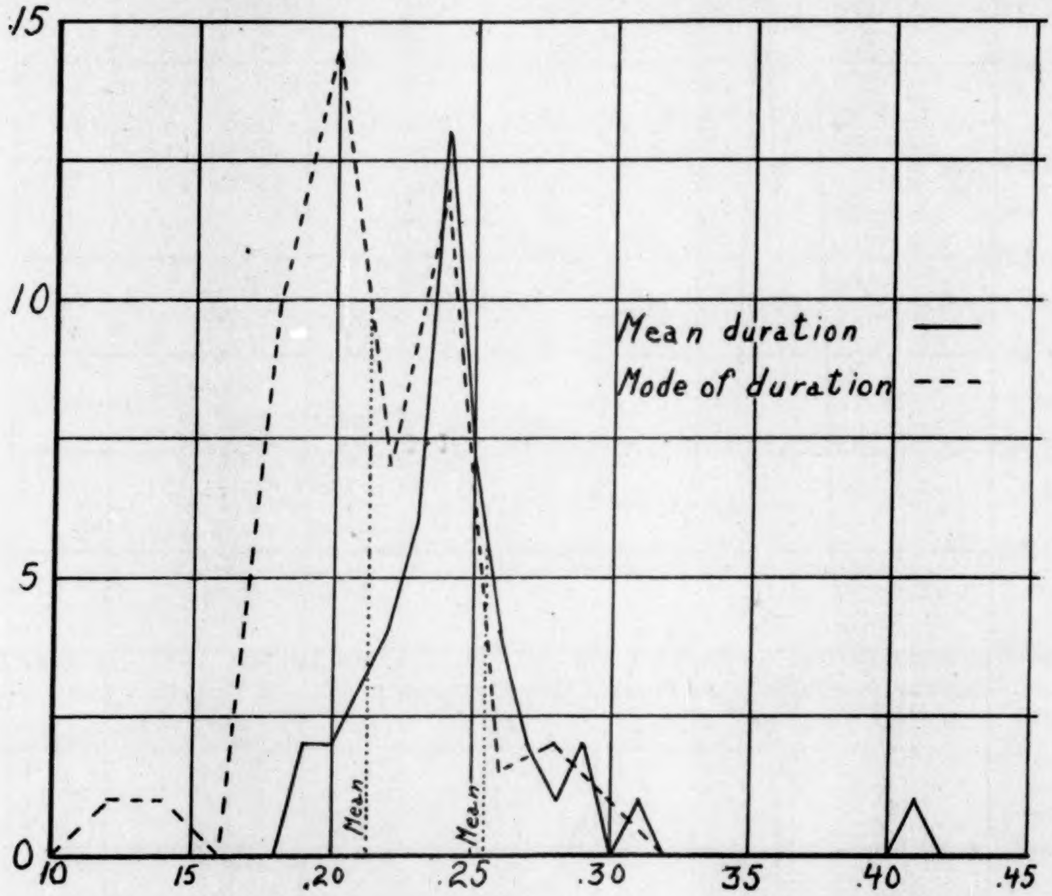

FIG. 3. The distribution curves for the means and modes of duration of the 50 good readers on Selection II. Notation same as in Fig. 2.

the duration of fixation could not be reduced to any great extent, it might be very greatly increased. This factor would tend to increase the average duration time.

Such a condition was found to exist. The mode of the duration fell at .213 sec. as compared to .255 sec., or .04 sec. shorter than the mean duration. The distribution curves of mean and mode of duration are shown in Fig. 3.

Closely related to the duration of fixation is the average amount

of material observed at each fixation. Our group of good readers averaged 2.92 ems (1.30 words) per fixation; S.D., .6. On account of the differences in size of type and length of line, it was not possible to make a direct comparison between other studies and this one. *Tinker and Paterson* (25) point out that, within moderate limits, change in size of type has very little effect upon the number of fixations per line or upon the reading rate.

Perception-time. Total perception-time is the product of the number of fixations and the average duration per fixation. This method discounts the interfixation-time, or time of movement. Table III gives the values for perception-time and interfixation-time for some of the earlier studies and for this study. It will be noted that results for this study indicate a rather pronounced difference, with perception-time averaging only 90 per cent of the total time. This can probably be best accounted for by the high rate of reading of this group. The interfixation-time can not be reduced to any great extent as it is chiefly physiological in nature. Several observers have pointed out the fact that the interfixation-movements are fairly constant in duration so that their time factor will remain constant.

The perception of the material occurs only during the fixation-time, with no perception occurring during movement of the eyes. *Erdmann and Dodge* (4) first pointed out that if a printed line passed by a window at the speed of 1 cm. in .01 sec., which has been found to be approximately the speed of saccadic eye-movements employed in reading, the material could not be perceived sufficiently well to be read. With an improved technique, *Dodge* (3) later showed that there is no clear vision during saccadic eye-movements.

TABLE III. *Per cent of the total time that is required for perception and interfixation-movements of the eye*

Experimenter	Material	Perception-time	Interfixation-time
Erdmann and Dodge	Prose	93.6	6.4
Shen	Chinese		
	Horizontal	94.4–96	4.0–5.6
	Vertical	94.5–95.9	4.1–5.6
Tinker	Prose	92.4	7.6
	Algebra	93.5	6.5
	Chemistry	93.4	6.6
	Formulae	90.	10.0
Walker	Prose	90.	10.0

Interfixation-movements. Since by far the majority of saccadic movements were less than .02 sec. and those of the return

sweep less than .04 sec. in duration, the unit of measurement employed with this camera does not permit reliable computation of these phenomena. *Tinker* (19) cites the speed of angular velocity of the eye-movements as calculated by various experimenters. The total interfixation-time for this group of good readers was 10 per cent. As pointed out in the paragraph on perception-time, this was a rather high figure, but an attempt has been made to account for the large deviation. The return sweep as a rule included approximately 70 to 95 per cent of the length of the line. There was no definite point for departure from the preceding line, nor an exact point for the beginning of the following line. The first fixation normally lay fairly close to the left end of the line, with not as great a deviation around this point as was the case with the last fixation in the line. Some cases showed a pronounced tendency toward failure to approximate this point, resulting in a second shift. This has been considered under regressions as Type 1. This defect is not so pronounced in the good reader as *Robinson* (16) found it to be in the poor reader.

The saccadic forward movements were found to be very regular in size: 4.31 ems; S.D., .65 ems. This standard deviation is equivalent to one-half of a letter, which showed that this particular movement was very constant. The size of the forward shift, 4.31 ems, did not agree with the average size of fixation, 2.92 ems. This discrepancy was accounted for by the regressions. As a rule the regressions tended to be greater than the forward shift, requiring several additional fixations to recover the material.

In records of eye-movements an occasional shift toward the left of the line is observed, which is not a shift to a succeeding line. These regressions are of some importance. If they are too numerous, they will have a very pronounced effect of decreasing the rate of reading. In this study two types of regression were found: (1) those following the first fixation in a line, due primarily to failure on the part of the reader to make a sufficiently large return sweep, or to an improper location in the line for the first fixation; (2) those occurring at any other point in the line. For this reason, the first regression is not considered in the statistical treatment, as it properly belongs to the return sweep. *Bus-*

well (1) has pointed out the fact that there is a tendency for regressions to occur more frequently at the first of the line, which he claims is due to lack of clearness of the initial part of the line. The second type of regression is probably caused by the reader wishing to check on a difficult phrase or word in order to clarify the meaning of the passage. There is no constancy in the amount of regression that may occur. In some cases, a reader may return to the first part of the word upon which he is fixating, or again, he may return a whole phrase, or almost to the first of the line. Disregarding the first type of regression and considering only Type 2, an average of .7 regressions per line was found for the good readers.

Rate of reading. Closely related to duration and frequency of fixation, and of most importance to the clinician, is the rate of reading. The reading rate of a moderately difficult selection for a group of good readers was found to be 11.51 ems (5.13 words) per sec., S.D., 1.84 ems. Expressed as ems per min., this group showed the rather rapid reading rate of 690 ems (308.03 words) per min. This was almost 224 ems (100 words) per min., or 50 per cent, faster than *Robinson* (15) found for a group of poor readers. Both groups of good and poor readers were selected from the same entering classes and evaluated on the same tests.

Minor phenomena. In some cases, where there were abnormalities of the adjusting mechanism of the eye, slight variations were observed. In a few of the subjects, there was at times an indication of a tremor, but the .02 sec. unit of time measurement did not permit accurate investigation. *Jasper and Murray* (9), working with stutterers, have shown that this tremor may be quite pronounced. The good readers failed to show any significant variations in the vertical movements, which *Jasper and Murray*(9) found to be quite pronounced in stutterers. In some of these cases evidence was found pointing to the existence of a dominant eye. This was shown by the dominant eye maintaining its point of regard, while the non-dominant eye, which had evidently converged or diverged during the shift, was required to make the correct accommodation during the early period of the fixation.

The time required for this accommodation varied from .02 to .06 sec.[4]

Head-movements occurred occasionally in the records of this study, but were very infrequent. In practically all cases, the experimenter was aware of the fact that the movement had occurred during the recording. When the subject was questioned, he attributed the movement to a tension, or to an uncomfortable position. *Gray* (6) pointed out that records which he had to discard due to head-movements were in every case those of poor readers. *Jasper and Murray* (9) also found very pronounced head-movement in the case of stutterers.

The dominance of one eye previously mentioned might be used as an index of laterality. *Jasper* (8) found that dominance of one eye, as measured by convergence and divergence, was one index of handedness. *Judd* (11), on the other hand, contended that this dissimilarity was not dependent upon right- or left-eyedness, or differences of visual acuity. He attributed it to a difference of muscular balance, because the degree of dissimilarity was changed with an alteration of the muscular tension brought about by a shift in the location of the fixation-points.

Effect of Difficulty of Reading Material on Eye-Movements

The second phase of this problem was the effect of variations of difficulty of reading material upon the major eye-movement phenomena. This was accomplished by analyzing records obtained from subjects while they were reading three selections of material of varying difficulty. Selection I, slightly under seven lines in length, was very simple elementary material; II, slightly under eleven lines in length, was of moderate difficulty for the average college student. Selection III, which was a paragraph

[4] One of the subjects selected for the study could not be used. In attempting to photograph his eye-movements, removal of his glasses apparently caused him to lose control of one eye, which wandered at random. Upon questioning, it was discovered that the subject was not aware of this abnormality, and it was very evident that he was using only one eye for reading, because he was able to maintain a very high rate. This particular subject had a percentile ranking of 100 for reading ability.

slightly over 17 lines in length, was taken from a scientific article of extreme difficulty. Comprehension was checked in the same manner as in Selection II. While the material in III might not have been at all familiar to the reader, care was taken to compose the questions for the check on comprehension so that no additional knowledge would be required of the reader.

Table I shows the results for this group of subjects on the various phenomena of eye-movements for the three selections of reading material used in this part of the study. The C.R. is shown in Table II for the differences between any two paragraphs for the different measures. The various eye-movement phenomena will be considered in the following pages in the same order as in the previous selection.

Fixations. The mean duration per fixation increased slightly with each increase of difficulty of reading material. The C.R. between I and II was not reliable; between II and III, and I and III, the difference was statistically quite significant, being 4.03 and 3.02. The obtained differences, .031 and .022 sec., were from a practical standpoint not so great, but the accumulation of this difference throughout the paragraph resulted in a significant difference.

A C.R. of 3.6 was found for the variability of the duration per fixation between I and III, which were the two extremes in difficulty of reading material. This difference was of no practical significance, for the obtained difference between any of the two selections was less than .02. This indicated that in respect to variability there was very little difference as a result of increase of difficulty of reading material.

No important difference was found for the mode of duration between the three paragraphs. A statistically reliable C.R. of 3.3 occurred between II and III, but the obtained difference was only .02 sec. which could not be considered of great importance.

Buswell (1) reported that the duration of fixation tended to remain constant after the sixth grade. *Tinker* (19) went further, claiming that there was a "habit-pause duration" tendency indicated by the constancy of the duration of fixation. *Miles and Segel* (14) claim that "the duration of fixation tends to remain the same for different lines and for different selections." *Judd and Buswell* (12), using three readers on five paragraphs ranked for increasing

difficulty in oral reading, also found a slight increase in duration of fixation. They found a difference when working with different types of reading material—fiction, geography, and rhetoric—but these differences occurred only within the individual. Unfortunately, they did not obtain introspections from their subjects as to the interest the particular paragraph might have had for the reader, so that the particular differences that occurred might have been due to the interest or knowledge of the individual reader.

With increasing difficulty of reading material, there was a notable decrease in the ems per fixation. The variability of the size of fixation also showed some difference. The C.R.'s between the selections were highly significant: for I and II, 3.65; for II and III, 3.23; and for I and III, 7.0. These particular figures do not mean that such an increase will necessarily hold for each individual. The coefficient of correlation for the number of ems per fixation for I and II was .33; for II and III, .40; and for I and III, .19. These figures plainly show that it would not be possible to predict from the results obtained on one paragraph what any individual would do on another paragraph.

Perception-time. The perception-time for each of the three selections was 90 per cent of the total time while interfixation-time was 10 per cent. No difference could be expected in these phases since the interfixation-time is largely dependent on physiological factors.

Interfixation-movement. The size of the forward shift changed very slightly with change in the difficulty of the reading material. Between I and II there was practically no difference in the size of the forward shift. Between I and III and II and III there was a statistically significant difference, but since the actual difference was not greater than one-quarter em, this measure can be considered as very constant.

The more difficult the material to be read, the greater the number of regressions. As pointed out earlier, regressions are due chiefly to lack of comprehension or to a desire on the part of the reader to recheck. Hence an increase in the number of regressions is to be expected. Selection I showed .44 regressions per line; II, .70 regressions; and III, .93 regressions. The variability was the same for I and II, .06 regressions per line, while

III showed a slightly greater variability with .12 regressions per line.

Rate of reading. Increase in difficulty of reading material caused a marked decrease in the rate of reading. This is commonly expected to be the result, for it is a matter of every-day experience that more difficult reading material takes greater time to read. Selection I was read at the exceptionally high rate of 757 ems (380 words) per min.; II, at the rate of 690 ems (308 words) per min.; and III, at 556 ems (242 words) per min. The variability among the subjects also decreased with increasing difficulty of material. The S.D. for I was 130 ems per min.; II, 110 ems, and III, 69 ems. The C.R. between any pair of selections indicates a high reliability for these differences.

Effect of Comprehension Requirements on Eye Movements

In this phase of the study, three selections of equal difficulty were presented to the subject. Special emphasis was placed on the requirements of comprehension of the paragraph to be read. Selection IV was studied for the purpose of obtaining records when only a general idea of the content was desired; II, of the preceding section, was used for moderate comprehension; and V, for detailed knowledge of the content. The subject was checked for comprehension on IV and V by having him give a detailed report to the experimenter at the conclusion of the reading.

C. T. Gray (6) found that there was a tendency to read in smaller units when the subject was required to reproduce the general idea of the paragraph. Answering questions on the material reduced the unit still further. *Judd and Buswell* (12) found a variety of reactions in readers when they were required to read a paragraph carefully in order to answer questions on the content. *Vernon* (24), in reviewing this study, contended that it is quite probable that very few of the subjects knew what correct procedure to adopt when faced with this problem. *Judd and Buswell* themselves stated that the subject had never been instructed as to the proper procedure to adopt under such circumstances. When the subjects were required to paraphrase the material, all semblance of normal reading disappeared with a great increase in the number of fixations and regressions.

Fixations. Increase in the requirement of comprehension tends to increase the duration of fixation. The C.R. between

any two of the group was highly reliable and, while the differences in means did not seem to be very great, the total difference for a paragraph would be quite large. There was a slightly greater obtained difference between IV and V than existed between IV and II.

The variability of the duration of the fixation also showed an increase with increasing requirements of comprehension. While there was no reliable C.R. between IV and II, a highly reliable C.R. existed between IV and V and II and V.

The mode of the duration of fixation showed very little difference between any two measures with change in the requirements of comprehension. No C.R. of any reliability existed between any of the selections. Again, the mode of duration was considerably smaller than the mean, which was further evidence of the fact that the range of duration per fixation was highly skewed, with a large grouping at the lower end of the distribution.

A subject necessarily reduces the size of the fixation with an increase in the requirements of comprehension. The average number of ems per fixation for IV, II, and V were 3.75, 2.92, and 2.52, respectively. These figures all gave a highly reliable C.R. between any two of the selections.

While on the face, these figures seem to support *Judd and Buswell's* (12) contention that readers tend to divide the material into smaller units when a more complete comprehension is required, a further study of the results fails to bear them out. Coefficients of correlation between any two measures were not reliable. They also contended that when the subject was confronted with more difficult material, he might either maintain the same duration per fixation and increase the size of fixation, or increase the duration and maintain the same size of fixation. If this were true, a scattergram of these two variables should show a bimodal grouping. To test this hypothesis, a scattergram of these two measures was plotted for the 50 subjects in this study. A normal unimodal curve was obtained.

Perception-time. The results here were similar to those obtained in the previous section of the study on varying degrees of

difficulty of reading material. Selection IV showed that 91 per cent of the total time was used for perception, and II and V, 90 per cent of the total time.

Interfixation-movements. The interfixation-time for IV was 9 per cent, and for II and V, 10 per cent of the total time. This is further evidence supporting the fact that interfixation-movements are probably habitual and physiological in nature.

There was a slight difference between the size of forward shift of any two of the three paragraphs used in this portion of the study. The mean forward shift for IV was 4.55 ems, for II, 4.31, and for V, 3.91. The C.R. between IV and II was 2.13, which was not highly reliable; between II and V, 3.13; and between IV and V, 4.77. The last two were highly reliable. The obtained difference between IV and V was .64 ems which, while reliably different, can be considered of little importance, since such slight changes will not total up to any great sum even for a large amount of reading material.

The regressions occurring in the body of the line showed a rather noticeable increase with increase in the requirement of comprehension. Large differences and highly reliable C.R.'s were found between any two paragraphs, indicating that increase in comprehension caused the subject to make more rechecks on the material previously read. Rather low coefficients of correlation were obtained between any two paragraphs.

Rate of reading. The present readers gave evidence of a progressive reduction in the rate of reading with increasing requirements of the knowledge of the contents of the paragraph read. The average rates for IV, II, and V were 882 (397 words), 690 (308 words), and 595 ems (287 words) per min., respectively. These differences are all reliable. The reductions in rate of reading were apparently the natural result of the changes noted in the size and duration of the fixations and were partially due to the changes in the frequency of regressions.

Skimming. As it is generally used in relation to reading, skimming seems to have no definite meaning. A reader is sometimes said to be skimming when merely glancing through an

article for a general idea of its contents. Again, he may be looking for a specific item such as a date, a name, or a formula. In this particular study, an attempt was made to isolate one type of skimming, *i.e.,* looking for a specific item. The reader was required to look for a passage in the paragraph which he would have to read in order to answer a question which had previonsly been asked of him.

In all measures, the records obtained from the subjects on this selection for the study of skimming were similar to those obtained from II. The results seem to contradict the general meaning of the word "skimming," but apply only to this definition of skimming. They warrant more detailed study of this aspect of reading, because other types of skimming would undoubtedly result in considerable differences for any or all of these measures. Comparison may be made between II and VI in Table I.

Relationship between span of perception and span of fixation

The data on the span of perception, as measured by word groups, in Table IV showed that the subjects were able to perceive units of 8.93 ems (3.94 words). These were much larger units of material than the fixation span of 4.31 ems (1.92 words) employed in normal reading. *Robinson* (**15**) working on poor readers found a span of perception of 2.92 words before training on size of fixation, and 3.03 after training.[5] *Robinson* (**15**) also showed that it is possible to increase the size of fixation in poor readers. Their average size of fixation in reading, 1.05 words before training, 1.70 after training, was smaller than the 1.92 of the good readers. From these data it is possible to expect that some improvement could be made in the reading span of fixation, but it would not approximate the span of perception. Hence it is questionable whether such training would result in an increase of rate for good readers. The possible improvement would hardly warrant the time and energy required for such training.

[5] The material used by *Robinson* (17) for his span of perception test was not comparable in difficulty to that used in this study. Reading material for both studies was comparable in difficulty and comprehension to Selection II.

TABLE IV. *Span of perception in ems as measured by span of perception cards and from records of reading*

Reading selection	I	II	III
Mean span as measured by cards...........	9.03	8.93	8.06
	1.76	2.13	2.45
Mean span as measured by records.........	4.31	4.31	4.06
	.75	.64	.56

Validity and reliability of objective methods

There is some question in the literature as to the validity of photographic technique of studying eye-movements. The cost and labor involved in this type of examination must always be considered if a great many tests are to be undertaken. *Miles* (13) contended that the use of the " peep-hole " method of observation was satisfactory for the ordinary clinical purpose. Such a method can only give an estimation of the gross movements and requires particular training on the part of the subject before accurate measurement can be made. To a certain extent the number of regressions and the number of fixations may be counted and the total time required for the reading of the passage may be measured. This technique, however, will not give the duration of the fixation, the size of the fixation, the location of the regressions, and probably none of the very small shifts. All of these factors enter into the total picture of reading rate and they should be considered in the analysis.

A check of the reliability of the two methods was carried out as a related phase of this study by counting all observable eye-movements during the photographic recording. After analysis of the record, intercorrelations were calculated on three factors entering in the frequency of fixation. Table V shows that the intercorrelations between the number of observed fixations, the number of calculated fixations, and the total time are all rather high, ranging from .90 ± .01 to .96 ± .01. When the partial correlation between observed fixations and calculated fixations was factored, holding the time element constant, the result was found to be .23. The calculated fixations were known to be accurate and not dependent upon subjective factors, therefore the low correlation between the two measures indicated that the observational method was not reliable. This was further evidence in support of

the same findings reported by *Robinson and Murphy* (16). If only a cursory examination is desired, the "peep-hole" method may be satisfactory.

To check the reliability of the measurements made by use of the pantograph, two records were reread for the amount of forward shift, no marks having been made on the records which would aid in rereading. The reliability between these two readings obtained by correlating the first with the second was .95 ± .01. Hence the method used was highly reliable in spite of the subjective factor present in the measurement. It will be noted that this was merely a measurement of the amount of forward progression and not a measurement of the location of the fixation within the line. The latter would undoubtedly be much less reliable.

TABLE V. *Correlations between the two methods of counting eye-movements, the total time for the passage, and the partial correlation between the two methods with time held constant*

Variable	Partial	.r	P.E.
1. Calculated fixations	r_{12}	.90	.01
2. Observed fixations	r_{13}	.96	.005
3. Total time	r_{23}	.91	.009
	$r_{12.3}$	.23	

Summary

For reading material of moderate difficulty and requirement of comprehension, 25 ems to the line, averaging 2.24 ems per word (Table I), the following measures, which constitute norms of eye-movements, were determined for good readers:

1. The mean duration per fixation was .255 sec.; S.D., .033; range, .155–.355 sec.

2. The mean of the individual S.D.'s of duration per fixation was .100 sec.; S.D., .032; range, .003–.198.

3. The mean mode of duration was .213 sec.; S.D., .033; range, .114–.312.

4. Due to the fact that the mean mode of duration was less than mean duration per fixation, the grouping did not fall along a normal curve, but was skewed to the lower end.

5. There were 2.92 ems per fixation; S.D., .68; range, 1.12–4.72.

6. The extent of shift for each forward fixation was 4.31 ems; S.D., .64; range, 1.39–6.23.

7. There were .70 regressions per line; S.D., .06; range, .51–.99. Only those regressions occurring in the body of the line were considered, because regressions that occurred as a second fixation in the line properly belong with the return sweep.

8. The reading rate was 691 ems (308 words) per minute; S.D., 110 ems (49 words); range, 358–1022 ems (160–456 words).

9. There were no significant minor variations among the individuals of this group.

10. There were no observable variations in the vertical eye-movements.

11. Few significant differences in movements were found between the two eyes.

For (1) increase in difficulty of reading material and (2) increase in the requirement of comprehension, the following facts (measures in Table I, C.R. in Table II) were established:

1. The duration per fixation increased with increase of difficulty of reading material. It also increased with increase of the requirement of comprehension. The mode of duration remained constant with these changes.

2. The size of fixation decreased with increase of difficulty of reading material. It also decreased with increase of the requirement of comprehension.

3. The rate of reading decreased with increase of difficulty of reading material. It also decreased with increase of the requirement of comprehension.

4. The extent of forward shift was fairly constant for changes in difficulty of reading material or of the requirement of comprehension.

5. The eye-movements for a paragraph of moderate difficulty and moderate comprehension were similar to the eye-movements of skimming, when skimming is defined as looking for a specific passage.

For the relationship between span of perception, measured by word groups and width of fixation in reading a paragraph of

moderate difficulty and comprehension, the following facts were established:

1. The span of perception was 8.93 ems; S.D., 2.13.
2. The width of fixations was 4.31 ems; S.D., .64.

The following conditions were noted for the photographic technique employed:

1. The correlation for reliability between this method and the direct observational method, with time held constant, was .23.
2. The unit of measurement, .02 sec., was not fine enough to permit the study of interfixation movement, but adequate for all other purposes.
3. The reliability of reading extent of forward shift in the line by means of a pantograph was .95 ± .01.
4. Since there were no variations in the vertical movements, it is not necessary to obtain these records when studying eye-movements and, since few differences in movements between the two eyes were noted, it is not necessary to obtain binocular records of good readers.

Bibliography

1. Buswell, G. T. Fundamental reading habits: a study of their development, *Suppl. Educ. Monog.*, No. 21, 1922.
2. Dearborn, W. F. The psychology of reading, *Columbia Univ. Contrib. to Philos. and Psychol.*, 14, No. 2, 1906.
3. Dodge, R. Visual perception during eye-movement, *Psychol. Rev.*, 7, 1900, 193–199.
4. Erdmann, B., and Dodge, R. Psychologische Untersuchungen über das Lesen auf experimenteller Grundlage, Halle, 1898.
5. Gilliland, A. R. Experimental studies of the effect on reading of changes in certain sensory factors (unpublished thesis), Univ. of Chicago, 1922.
6. Gray, C. T. Types of reading ability as exhibited through tests and laboratory experiments, *Suppl. Educ. Monog.*, No. 22, 1922.
7. Huey, E. B. On the psychology and physiology of reading II, *Amer. J. Psychol.*, 12, 1901, 292–312.
8. Jasper, H. H. A laboratory study of diagnostic indices of bilateral neuromuscular organization in stutterers and normal speakers, *Psychol. Monog.*, 43, 1932, No. 1, 72–174.
9. Jasper, H. H., and Murray, E. A study of the eye-movements of stutterers during oral reading, *J. Exper. Psychol.*, 15, 1932, 528–538.
10. Jasper, H. H., and Walker, R. Y. The Iowa eye-movement camera, *Science*, 74, 1931, 291–294.
11. Judd, C. H. Photographic records of convergence and divergence, *Yale Psychol. Studies*, N.S., I, No. 2.
12. Judd, C. H., and Buswell, G. T. Silent reading: a study of the various types, *Suppl. Educ. Monog.*, No. 23, 1922.

13. MILES, W. R. The peep-hole method for observing eye-movements in reading, *J. Gen. Psychol.*, 1, 1927, 373–374.
14. MILES, W. R., and SEGEL, D. Clinical observation on eye-movements in the rating of reading ability, *J. Educ. Psychol.*, 20, 1929, 520–529.
15. ROBINSON, F. P. The rôle of eye-movements in reading with an evaluation of techniques for their improvement, *Univ. of Iowa: Series on Aims and Progress of Research*, No. 39, 1933.
16. ROBINSON, F. P., and MURPHY, P. G. The validity of measuring eye-movements by direct observation, *Science*, 76, 1932, 171–172.
17. SCHMIDT, W. A. An experimental study in the psychology of reading, *Suppl. Educ. Monog.*, 1, 1917, No. 2.
18. SHEN, E. An analysis of eye-movements in the reading of Chinese, *J. Exper. Psychol.*, 10, 1927, 158–183.
19. TINKER, M. A. Eye-movement duration, pause duration, and reading time, *Psychol. Rev.*, 35, 1928, 385–397.
20. TINKER, M. A. A photographic study of eye-movements in reading formulae, *Genet. Psychol. Monog.*, 3, 1928, 68–182.
21. TINKER, M. A. Visual apprehension and perception in reading, *Psychol. Bull.*, 26, 1929, 223–240.
22. TINKER, M. A., and PATERSON, D. G. Influence of type form on speed of reading, *J. Appl. Psychol.*, 12, 1928, 359–368.
23. TINKER, M. A., and PATERSON, D. G. Studies of typographical factors influencing speed of reading: III. Length of line, *J. Appl. Psychol.*, 13, 1929, 205–219.
24. VERNON, M. D. The experimental study of reading. London, Cambridge Univ. Press, 1931.

Contents of numbers VI-XVI

UNIVERSITY OF IOWA STUDIES IN PSYCHOLOGY

VI. 1914 (Price, $1.75)

THE TONOSCOPE.......................................Carl E. Seashore 1
ACCURACY OF THE VOICE IN SIMPLE PITCH SINGING......Walter R. Miles 13
THE EFFECT OF TRAINING IN PITCH DISCRIMINATION
Franklin Orion Smith 67
THE LOWER LIMIT OF TONALITY................Thomas Franklin Vance 104
VARIATION IN PITCH DISCRIMINATION..........Thomas Franklin Vance 115
THE DURATION OF TONES........................David Allen Anderson 150
THE EFFECT OF INTENSITY AND ORDER...............Rolland M. Stewart 157
THE EFFECT OF THE INTENSITY OF SOUND................Clara Hancock 161
THE MEASUREMENT OF TIME SENSE..................Felix Bruene Ross 166
SOME STANDARDIZING TESTS......................Reuel H. Sylvester 173

VII. 1918 (Price, $1.75)

DESCRIPTION OF AN UNUSUAL CASE OF PARTIAL COLOR BLINDNESS
Mabel Clare Williams 1
BINAURAL BEATS...G. W. Stewart 31
CORRELATION OF FACTORS IN MUSICAL TALENT AND TRAINING
C. E. Seashore and George H. Mount 31
THE PERCEPTION OF CONSONANCE AND DISSONANCE......C. F. Malmberg 93
A REVISION OF THE CONSONANCE TEST...............Esther Allen Gaw 134
THE COMPARATIVE SENSITIVENESS OF BLIND AND SEEING PERSONS
C. E. Seashore and T. L. Ling 148
THE ELEMENTAL CHARACTER OF SENSORY DISCRIMINATION
C. E. Seashore and Kwei Tan 159

VIII. 1922 (Price, $4.00)

WAVE PHASE IN THE OPEN-AIR LOCALIZATION OF SOUND..C. E. Seashore 1
THE RÔLE OF INTENSITY IN AUDITORY WAVE PHASE..Henry M. Halverson 7
THE INTENSITY LOGARITHMIC LAW AND THE DIFFERENCE OF PHASE EFFECT IN BINAURAL AUDITION.............................G. W. Stewart 30
MEASUREMENT OF ACUITY OF HEARING THROUGHOUT THE TONAL RANGE
Cordia C. Bunch 45
MEASUREMENT OF AUDITORY ACUITY WITH THE IOWA PITCH RANGE AUDIOMETER..............................Benjamin Franklin Zuehl 83
A STROBOSCOPIC DEVICE FOR MEASURING REVOLUTION RATES
Benjamin Franklin Zuehl 98
VISUAL TRAINING OF THE PITCH OF THE VOICE............Carl J. Knock 102
A SURVEY OF MUSICAL TALENT IN A MUSIC SCHOOL....Esther Allen Gaw 128
THE INHERITANCE OF SPECIFIC MUSICAL CAPACITIES
Hazel Martha Stanton 157
VOICE INFLECTION IN SPEECH..........................Glenn N. Merry 205
AN EXPERIMENTAL STUDY OF THE PITCH FACTOR IN ARTISTIC SINGING
Max Schoen 231
VOLUNTARY CONTROL OF THE INTENSITY OF SOUND
Dorothea Emeline Wickham 260
A COMPARISON OF AUDITORY IMAGES OF MUSICIANS, PSYCHOLOGISTS AND CHILDREN...Marie Agnew 268
THE AUDITORY IMAGERY OF GREAT COMPOSERS............Marie Agnew 279

A Pursuit Apparatus: Eye-Hand Coördination...Wilhelmine Koerth 288
The Tapping Test: A Measure of Motility..........Merrill J. Ream 293
Serial Action as a Basic Measure of Motor Capacity
C. Frederick Hansen 320

IX. 1926 (Price, $3.00)

Technique for Objective Studies of the Vocal Art..Milton Metfessel 1
The Variability of Consecutive Wave-Lengths in Vocal and Instrumental Sounds..................................Clarence Simon 41
The Vibrato.....................................Jacob Kwalwasser 84
A Phono-Photographic Study of the Stutterer's Voice and Speech
Lee Edward Travis 109
Studies in Motor Rhythm...................Robert Holmes Seashore 142
The Relation Between Musical Capacity and Performance
Flora Mercer Brennan 190
A Report of Three Singing Tests Given on the Tonoscope
Flora Mercer Brennan 249

X. 1926 (Price, $1.75)

A Phenomenon in Vision Similar to Refractory Phase
Roland C. Travis 1
The Diagnosis of Character Types by Visual and Auditory Thresholds....................................Roland C. Travis 18
An Experimental Study of Certain Tests as Measures of Natural Capacity and Aptitude for Typewriting....Benjamin W. Robinson 38
The Measurement of Capacity for Skill in Stenography
Oliver A. Ohmann 54
The Relation Between Faulty Speech and Lack of Certain Musical Talents.................Lee Edward Travis and Mildred G. Davis 71
The Basic Factors in the Human Voice...........Carl I. Erickson 82

XI. 1928 (Price, $3.50)

A Study of Fear by Means of the Psychogalvanic Technique
Nancy Bayley 1
The Application of Phi-Phenomena to Beats......Ingvald B. Hauge 39
An Experimental Critique of the Seashore Consonance Test
Delia L. Larson 49
The Mode of Vibration of the Vocal Cords.......Wolfgang Metzger 82
A Vacuum Tube Technique for the Frequency Range of Phase Effect.......................................Erving N. Peterson 160
Some Temporal Aspects of Sound Localization......Otis C. Trimble 172
A Note in Regard to the Extent of the Vibrato..Harold M. Williams 226

XII. 1928 (Price, $3.30)
SEASHORE COMMEMORATIVE NUMBER

Dean Seashore (Introduction).. 1
A Complete Annotated Bibliography of the Writings of Carl Emil Seashore........................J. E. Bathurst and R. D. Sinclair 3
Ten Volumes of Iowa Studies in Psychology Reviewed
Mabel Clare Williams Kemmerer 23
A Comparative Study of the Performances of Stutterers and Normal Speakers in Mirror Tracing.........Lee Edward Travis 45
Stanford Motor Skills Unit.................Robert Holmes Seashore 51
The Sense of Direction in Mental Imagery..........Carl I. Erickson 67
The Iowa State College Reasoning Test...........Thomas F. Vance 72
Vestibular Sensitivity to Intermittent Passive Rotation of the Body...Roland C. Travis 78
Iowa Placement Examinations—A New Departure in Mental Measurement.............................George D. Stoddard 92

Seashore's Plan of Sectioning on the Basis of Ability as a Motivation Device.......H. J. Arnold 102
Successful Teaching.......Russell Warrick Tallman 106
Some Aspects to be Considered in the Organization of a First Course in Psychology.......Alvhh R. Lauer 110
The Determination of a General Factor in Research Ability of College Students.......Franklin O. Smith 119
What is the Voice Vibrato?.......Milton Metfessel 126
Seashore Measures of Musical Talent.......Hazel M. Stanton 135
Five Studies of the Music Tests.......Esther Allen Gaw 145
Correlation Between Intelligence and Musical Talent Among University Students..George Cutler Fracker and Virgie M. Howard 157
The Aesthetic Attitude in Music.......Max Schoen 162
A Measure of Art Talent.......Norman C. Meier 184
The First Vocal Vibrations in the Attack in Singing
F. A. Stevens and W. R. Miles 200

XIII. 1930 (Price, $3.25)

Disturbances in Breathing during Stuttering.....Harold R. Fossler 1
Measurement in Musical Talent for the Prediction of Success in Instrumental Music.......William S. Larson 33
A Study in the Seashore Motor-Rhythm Test.......J. T. Nielsen 74
A Qualitative and Quantitative Study of the Emotion of Surprise
Emily Patterson 85
Pitch-Patterns and Tonal Movement in Speech......Alfred R. Root 109
An Experimental Study in Control of the Vocal Vibrato
Arnold H. Wagner 160

XIV. 1931 (Price, $4.00)

Voluntary Movements of the Organs of Speech in Stutterers and Non-Stutterers.......Bruce Blackburn 1
The Comparative Study of Affective Responses by Means of the Impressive and Expressive Methods.......Donald W. Dysinger 14
An Investigation of the Phenomena Connected with the Beating Complex.......Ingvald B. Hauge 32
Apparent Movement in Auditory Perception.......Anna Mathiesen 74
A "Central" Explanation of Sound Localization
Christian A. Ruckmick 132
Variations in the Galvanic Response.......Franklin O. Smith 142
The Psychophysics of the Vibrato.......Joseph Tiffin 153
The Relation Between the Physical Pattern and the Reproduction of Short Temporal Intervals.......Dorothy Triplett 201
Experimental Studies in the Use of the Tonoscope
Harold M. Williams 266

XV. 1932 (Price, $3.75)

A Phonophotographic Analysis of the Vocal Disturbances in Stuttering.......Bryng Bryngelson 1
An Action-Current and Reflex-Time Study of Psychiatric and Neurologic Cases.......Donald W. Dysinger 31
A Clinico-Experimental Approach to the Re-Education of the Speech of Stutterers.......Leo B. Fagan 53
Graphic Stuttering.......Leo B. Fagan 67
A Laboratory Study of Diagnostic Indices of Bilateral Neuromuscular Organization in Stutterers and Normal Speakers
Herbert H. Jasper 72
Some Common Factors in Reading and Speech Disabilities
George A. Kelly 175
The Remnant Capacities of the Feebleminded...William E. McClure 202

Dysintegration of Breathing and Eye-Movements in Stutterers During Silent Reading and Reasoning .... Elwood Murray 218
Causes of Maladjustment in Children .... Herbert D. Williams 276

XVI. 1933 (Price, $1.50)

The Effect of Pain and Emotional Stimuli and Alcohol upon Pupillary Reflex Activity .... W. Ralph Griggs Bender 1
Some Neuro-Physiological Sources of Action-Current Frequencies Donald B. Lindsley 33
A Comparative Study of the Achilles and the Patellar Reflex Response Latencies as Measured by the Action-Current and the Muscle-Thickening Methods .... Clarence L. Nystrom 61